Direct Sum and Partitionability Testing over General Groups

Andrej Bogdanov[*] Gautam Prakriya[†]

Abstract

A function $f(x_1,\ldots,x_n)$ from a product domain $\mathcal{D}_1\times\cdots\times\mathcal{D}_n$ to an abelian group $\mathcal{G}$ is a *direct sum* if it is of the form $f_1(x_1)+\cdots+f_n(x_n)$. We present a new 4-query direct sum test with optimal (up to constant factors) soundness error. This generalizes a result of Dinur and Golubev (RANDOM 2019) which is tailored to the target group $\mathcal{G}=\mathbb{Z}_2$. As a special case, we obtain an optimal *affinity test* for $\mathcal{G}$-valued functions on domain $\{0,1\}^n$ under product measure. Our analysis relies on the hypercontractivity of the binary erasure channel.

We also study the testability of *function partitionability* over product domains into disjoint components. A $\mathcal{G}$-valued $f(x_1,\ldots,x_n)$ is *k-direct sum partitionable* if it can be written as a sum of functions over k nonempty disjoint sets of inputs. A function $f(x_1,\ldots,x_n)$ with unstructured product range $\mathcal{R}^k$ is *direct product partitionable* if its outputs depend on disjoint sets of inputs.

We show that direct sum partitionability and direct product partitionability are one-sided error testable with $O((n-k)(\log n+1/\epsilon)+1/\epsilon)$ adaptive queries and $O((n/\epsilon)\log^2(n/\epsilon))$ non-adaptive queries, respectively. Both bounds are tight up to the logarithmic factors for constant ϵ even with respect to adaptive, two-sided error testers. We also give a non-adaptive one-sided error tester for direct sum partitionability with query complexity $O(kn^2(\log n)^2/\epsilon)$.

1 Introduction

In their seminal result, Blum, Luby and Rubinfeld [BLR90] gave a four query test to determine whether a function $f:\mathbb{F}_2^n\to\mathbb{F}_2$ is affine. We consider a natural generalization of the notion of affinity to functions $f(x_1,\cdots,x_n)$ from $\{0,1\}^n$ to an arbitrary abelian group $\mathcal{G}$: Is f of the form $x_1\cdot g_1+\cdots+x_n\cdot g_n+g_0$ for some group elements $g_0,\ldots,g_n\in\mathcal{G}$? The analysis of Blum, Luby and Rubinfeld does not apply unless there is a group homomorphism from the domain to the range.

In this work we give an optimal four query affinity test for functions from $\{0,1\}^n$ to an arbitrary abelian group $\mathcal{G}$.

More generally, our test can be used to determine if a function $f(x_1,\ldots,x_n)$ from a finite product domain $\mathcal{D}_1\times\cdots\times\mathcal{D}_n$ to an abelian group $\mathcal{G}$ is a *direct sum*, i.e., whether f is of the form $\sum f_i(x_i)$. This resolves a conjecture of Dinur and Golubev [DG19].

In contrast to the work of Blum, Luby, and Rubinfeld, which was primarily motivated by applications to probabilistically checkable proofs, direct sum testing over general groups arises in the context of testing *function partionability*: Can a multivariate function be decomposed into independent or loosely related components? Bogdanov and Wang [BW20] discuss the relevance

[*] andrejb@cse.cuhk.edu.hk. Department of Computer Science and Engineering and Institute of Theoretical Computer Science and Communications, Chinese University of Hong Kong.

[†] gautamprakriya@gmail.com. Institute of Theoretical Computer Science and Communications, Chinese University of Hong Kong.

ISSN 1433-8092

of this question for real-valued functions to the problem of identifying decompositions of control variables in high-dimensional reinforcement learning. In that setting a direct sum decomposition of the advantage function f describes a system that can be partitioned into independent components, which are lower-dimensional and therefore typically easier to learn. An efficient testing algorithm can be used to probe the existence of such a decomposition before any effort is expended into learning it.

In this work we consider the following two natural partitioning problems for discrete functions over product domains $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ (endowed with a product distribution):

- A *direct sum partition* ($\oplus$-partition) of f into k components is a representation of the form $f(x_1, \ldots, x_n) = f_1(x_{S_1}) + \cdots + f_k(x_{S_k})$, where $S_1, \ldots, S_k$ are disjoint nonempty sets of variables. Here, the range of f is an abelian group $(\mathcal{G}, +)$.
- A *direct product partition* ($\otimes$-partition) of f is a representation of the form $f(x_1, \ldots, x_n) = (f_1(x_{S_1}), \ldots, f_k(x_{S_k}))$, where $S_1, \ldots, S_k$ are disjoint nonempty sets of variables. Here, the range of f is a k-product set $\mathcal{R}^k$.

We are interested in the query complexity of testing partitionability: Given oracle access to f and parameters k, ϵ, how many queries does it take to tell whether f is partitionable or ϵ-far from partitionable?

The related tasks of *direct product testing* and *direct sum testing* ask for the existence of such representations under a known (fixed) partition of inputs. Motivated by applications to probabilistically checkable proofs, Dinur and Steurer [DS14] and Dickstein and Dinur [DD19] analyze a 2-query direct product test of essentially optimal soundness.

The query complexity of direct sum testing for $\mathbb{Z}_2$-valued functions, that is of testing whether a function $f\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathbb{Z}_2$ is of the form $f(x_1, \ldots, x_n) = f_1(x_1) + \cdots + f_n(x_n)$, was recently resolved by Dinur and Golubev [DG19]. They proposed and analysed a 4-query test of optimal (up to constant factors) soundness error. Their tester does not naturally extend to functions valued in arbitrary abelian groups.

Bogdanov and Wang [BW20] proposed an agnostic learning algorithm for unknown direct sum partitions. As a consequence of their analysis they concluded that $\oplus$-partitionability is testable with $O(kn^3/\epsilon)$ non-adaptive queries. They also showed that $\Omega(n - k + 1)$ queries are necessary for constant ϵ. To the best of our knowledge $\otimes$-partitionability has not been studied before.

Our Results

We analyze a new 4-query direct sum test for functions valued over arbitrary abelian groups. The test is based on the following dual characterization: $f\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$ is a direct sum $f_1(x_1) + \cdots + f_n(x_n)$ if and only if $D_f(S, \overline{S}; x, y) = 0$ for all pairs of inputs x, y and partitions $(S, \overline{S})$ of $[n]$, where

$$D_f(S, \overline{S}; x, y) = f(x) - f(y_S x) - f(y_{\overline{S}} x) + f(y).$$

Here and in the rest of the manuscript, $y_S x$ is the string in $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ that matches y in the S-coordinates and x in the other coordinates. We assume that the domain $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ is furnished with a product distribution: For x chosen at random from $\mathcal{D}_1 \times \cdots, \times \mathcal{D}_n$, the coordinates $x_1, \ldots, x_n$ are independent.

The tester accepts if $D_f(S, \overline{S}; x, y) = 0$ for random independent inputs $x, y \in \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ and a uniformly random partition $(S, \overline{S})$ of $[n]$. Our main result is an optimal (up to constant factor) bound on the soundness error $\rho(f) = \Pr[D_f(S, \overline{S}; x, y) \neq 0]$ of this test in terms of the distance $\delta(f) = \min_g\{\Pr_x[f(x) \neq g(x)] \colon g \text{ is a direct sum}\}$.

Theorem 1.1. *There is an absolute constant $c > 0$ such that for every collection of finite sets $\mathcal{D}_1, \dots, \mathcal{D}_n$, every abelian group $\mathcal{G}$, and every $f \colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$, $\rho(f) \geq c \cdot \delta(f)$.*

An important special case of the theorem concerns the Boolean domain $\mathcal{D}_1 = \cdots = \mathcal{D}_n = \{0, 1\}$ under the uniform distribution. The class of direct sums from $\{0, 1\}^n$ to $\mathcal{G}$ is then precisely the class of affine functions $f(x) = x_1g_1 + \cdots + x_ng_n + g_0$ for some group elements $g_0, g_1, \dots, g_n \in \mathcal{G}$ (see Claim 4.2).

Using Theorem 1.1, we obtain the following upper bound on the query complexity of $\oplus$-partitionability.

Theorem 1.2. *Direct sum partitionability over any abelian group is one-sided testable with $O((n-k)(\log n + 1/\epsilon) + 1/\epsilon)$-queries.*

We also prove an upper bound on the query complexity of $\otimes$-partitionability:

Theorem 1.3. *Direct product partitionability is one-sided testable with $O((n/\epsilon)\log^2(n/\epsilon))$ non-adaptive queries.*

The testers in Theorem 1.2 and 1.3 are time-efficient.

By the lower bound of Bogdanov and Wang, the $\oplus$-partitionability tester is tight up to the $\log n$ factor for constant ϵ. In the special case when $k = n$, direct sum partitionability reduces to direct sum testing and the query complexity is the same as that of Theorem 1.1.

Our tester for $\oplus$-partitionability is adaptive. In Theorem 5.6 we also give a one-sided non-adaptive tester of query complexity $O(kn^2(\log n)^2/\epsilon)$. A non-adaptive lower bound of $\Omega((n - k + 1)\,(\log(n - k + 1)/\epsilon^c)\log(\log(n - k + 1)/\epsilon^c))$ for any $c > 1$ follows from the work of Servedio et al. [STW15] on junta testing (see our Section 5.2). As in the case of juntas, it follows that adaptivity helps in testing $\oplus$-partitionability for some settings of parameters.

The $\otimes$-partitionability tester is also nearly tight: In Propositions 6.6 and 6.10 we show that direct product partitionability requires $\Omega(n)$ queries for $\epsilon = 1/2$ for adaptive testers, and $\Omega(\frac{n}{\epsilon \log(1/\epsilon)})$ queries for non-adaptive testers, for every $k \geq 2$.

Ideas and Techniques

Direct sum testing over general groups The main ingredient of Dinur and Golubev's direct sum tester for $\mathbb{Z}_2$-valued functions is an implicit reduction from general product domains to the Boolean domain $\{0, 1\}^n$ under the uniform distribution. We abstract and generalize their reduction in Proposition 4.1. To complete their proof, Dinur and Golubev instantiate the reduction with the $\mathbb{Z}_2$-affinity test of Blum, Luby, and Rubinfeld [BLR90].

Our main technical contribution is a tight analysis of the affinity test D_f applied to functions $f \colon \{0, 1\}^n \to \mathcal{G}$ valued in an arbitrary abelian group $\mathcal{G}$. To give a sense why the test is sound, let us argue that $\rho(f) = \Omega(\delta(f))$ under the additional assumption that f is close to a direct sum, say if $\delta = \delta(f) \leq 1/27$.

Let B be the set of measure at most $1/27$ on which f differs from its closest direct sum. We claim that conditioned on $x \in B$, the probability that any of the other test queries y, y_Sx, $y_{\overline{S}}x$

Property	Our Results	Prior Work
Direct Sum: $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$ s.t. $f(x) = f(x_1) + \cdots + f(x_n)$	4-query test for arbitrary abelian group $\mathcal{G}$ (Theorem 1.1)	4-query test for $\mathcal{G} = \mathbb{Z}_2$ [DG19]
k-$\oplus$-partitionability: $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$ s.t. $\exists S_1, \ldots, S_k \subseteq [n]$, $f(x) = f(x_{S_1}) + \cdots + f(x_{S_k})$	$O((n-k)(\log n + 1/\epsilon) + 1/\epsilon)$-query adaptive test (Theorem 1.2) $O(kn^2(\log n)^2/\epsilon)$-query non-adaptive test (Theorem 5.6)	$O(kn^3/\epsilon)$-query non-adaptive test [BW20]
k-$\otimes$-partitionability: $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{R}^k$ s.t. $\exists S_1, \ldots, S_k \subseteq [n]$, $f(x) = (f(x_{S_1}), \ldots, f(x_{S_k}))$	$O((n/\epsilon)\log^2(n/\epsilon))$-query non-adaptive test (Theorem 1.3)	

Table 1: Summary of algorithmic results. All tests have one-sided error.

land in B is at most $\delta + 2\delta^{1/3}$. By independence, the probability that $y \in B$ conditioned on $x \in B$ is exactly δ. In contrast, $y_S x$ and $y_{\overline{S}} x$ are not independent of x, but can be sampled by processing x through a binary symmetric channel with crossover probability $1/4$. The bound $\Pr[y_S x \in B | x \in B] \leq \delta^{1/3}$ follows from the small-set expansion of this channel [AG76], which is equivalent to the hypercontractivity of the corresponding Markov operator [Bon70]. Since the event "$x \in B$ and $y_S x \notin B$ and $y_{\overline{S}} x \notin B$ and $y \notin B$" results in rejection, it follows that $\rho(f) \geq \delta \cdot (1 - \delta - 2\delta^{1/3})$, which is at least $\frac{8}{27}\delta$ by the closeness assumption on f.

For larger values of $\delta(f)$, our proof strategy is to argue that f can be *decoded* to a direct sum function by making at most $O(\rho(f))$ "changes" to the truth-table of f. The decoding algorithm we analyze in Lemma 2.3 is *iterative plurality* (i.e., iterative maximum likelihood). We show that the function

$$\phi(x) = \text{plurality}_{S,y}\ f(y_S x) + f(y_{\overline{S}} x) - f(y) \tag{1}$$

is, on the one hand, $2\delta(f)$-close to f, and on the other hand, has substantially smaller rejection probability of $\rho(\phi) \leq \rho(f)/2$. By iterating the decoding, i.e. applying the plurality to ϕ again, we arrive at a function that is $4\delta(f)$ close to f and passes the test with probability one, thus must equal a direct product.

This argument is inspired by the linearity test analysis of Blum, Ruby, and Rubinfeld (BLR), who also decode f to a function that is, on the one hand, close to f and, on the other hand, passes their test with probability 1. However, unlike the BLR decoder which yields a linear function after a single round of self-correction, ours inherently requires multiple iterations. For example, if f is a direct sum corrupted on all inputs with relative hamming weight around $1/4$, then $\phi(0)$ is unlikely to be correctly decoded (as $y_S 0$ and $y_{\overline{S}} 0$ will typically be corrupted) and so will typically be inconsistent with a direct sum.

Nevertheless, the high-level structure of our argument closely parallels the BLR analysis. First, in Claim 3.4 we show that for all but an $o(\rho)$-fraction of inputs x, the plurality in (1) is a strong majority consistent with 99% of the choices of $(S, \overline{S})$ and y. Second, we use the algebraic structure of our test (Claim 3.3) to show that if $D_\phi(S, \overline{S}; x, y) \neq 0$ then $D_f(U, V; w, z) \neq 0$ for a substantially larger fraction of query sequences $(w, z_U w, z_{\overline{U}} w, z)$ that can be sampled by applying suitable "noise" to $(S, \overline{S}; x, y)$. If we represent the partition $(S, \overline{S})$ by a binary string $\sigma \in \{0,1\}^n$ (with 1 and 0 indicating memberships in S and $\overline{S}$, respectively), we show that the relevant noise can be modeled by independent fixed-probability *erasures* applied to the symbols of σ, x, and y. Using hypercontractivity bounds for the binary erasure channel [NW16], we conclude that ϕ fails the test on a significantly smaller fraction of queries than f does.

In the special case when the target group is $\mathbb{Z}_2$, the soundness error of D_f can be directly shown to be within a constant factor of the soundness error of the Dinur-Golubev tester (even though the two tests are different). The main motivating applications for function partitionability, however, concern real-valued functions [BW20]. The analysis of our $\oplus$-partionability testers for such functions relies on Theorem 1.1.

The idea of soundness analysis by iterative plurality decoding was introduced by Ben-Sasson et al. [BSSVW03] and used by Shpilka and Wigderson [SW06] in the context of randomness-efficient linearity testing.

Testing partitionability The main ingredient in our $\oplus$-partitionability algorithms is the direct 2-sum test D_f. The structure of this test allows us to efficiently detect a pair of variables x_s, x_t that must fall in the same component of the partition in any far from $\oplus$-partitionable function, effectively reducing the instance size by one variable.

Our $\otimes$-partitionability test looks for an input variable that is influential in at least two of the output coordinates of f. The analysis of this test is based on Lemma 6.4, which states that such a variable must exist in any far from partitionable function.

Organization

Section 2 outlines the proof of Theorem 1.1 in the case when the domain is the Boolean hypercube. The analysis is based on the convergence of the iterative decoder (Lemma 2.3), which is proved in Section 3. Section 4 analyzes the reduction from testing functions over arbitrary product domains to testing functions on the hypercube and proves Theorem 1.1. Sections 5 and 6 describe and analyze the partitionability testers for direct sum and direct product, respectively.

Definitions and Notation

Let $\mathcal{D} \doteq \mathcal{D}_1 \times \ldots \times \mathcal{D}_n$ be a finite set. For strings $x, y \in \mathcal{D}$ and a set of indices $S \subseteq [n]$, let x_S to refer to the projection of x onto the coordinates in S. For strings $x^{(1)}, \ldots x^{(k)} \in \mathcal{D}$, and a partition $S_1, \ldots, S_k$ of $[n]$, let $x^{(1)}_{S_1} \ldots x^{(k)}_{S_k}$ be the string in $\mathcal{D}$ that is identical to $x^{(i)}$ on indices in S_i. For a bipartition $(S, \overline{S})$, we often write $x_S y$ instead of $x_S y_{\overline{S}}$.

In sections 2 and 3 we identify a bipartition $(S, \overline{S})$ of $[n]$ with its indicator vector $\sigma \in \{0,1\}^n$, and write $D_f(\sigma; x, y)$ instead of $D_f(S, \overline{S}; x, y)$, and x_σ instead of x_S.

We extend the definition of D_f to pairs of disjoint sets (S, T) that do not necessarily partition $[n]$ as

$$D_f(S, T; x, y) \doteq f(x) - f(y_S x) - f(y_T x) + f(y_{S\cup T} x).$$

2 Direct Sum Test for Functions on the Boolean Hypercube

The following dual characterization of direct sums motivates our test.

Fact 2.1. *A function $f : \{0,1\}^n \to \mathcal{G}$ is a direct sum if and only if $D_f(\pi; x, y) = 0$ for every choice of $x, y, \pi \in \{0,1\}^n$.*

Proof of Fact 2.1. The 'only if' direction is immediate from the definition of a direct sum. We prove the 'if' direction. Let f be such that $D_f(\pi; x, y) = 0$ for every choice of $x, y, \pi \in \{0,1\}^n$. Fix $y \in \{0,1\}^n$. For every $x \in \{0,1\}^n$ we can write $f(x)$ as

$$\begin{aligned} f(x) &= f(x_{\{1\}}y) + f(y_{\{1\}}x) - f(y) \\ &= f(x_{\{1\}}y) + f(x_{\{2\}}y) + f(y_{\{1,2\}}x) - 2f(y) \\ &\vdots \\ &= f(x_{\{1\}}y) + f(x_{\{2\}}y) + \ldots + f(x_{\{n\}}y) - (n-1)f(y). \end{aligned}$$

Therefore, f is a direct sum. □

```
Algorithm 1: Direct sum test for functions over $\{0,1\}^n$
  Oracle : $f\colon \{0,1\}^n \to \mathcal{G}$
  Sample $x, y, \pi \in \{0,1\}^n$ independently and uniformly at random.
  If $f(x) + f(y) - f(x_\pi y) - f(y_\pi x) = 0$, accept.
  Else, reject.
```

By Fact 2.1, the test accepts every direct sum with probability 1. The following proposition establishes soundness of the test. Let $\rho(f)$ denote the probability that Algorithm 1 rejects the function f. That is, $\rho(f) \doteq \Pr_{x,y,\pi}[D_f(\pi; x, y) \neq 0]$.

Proposition 2.2 (Soundness)**.** *There exist a universal constant $\eta \in [0,1]$ such that for every function $f : \{0,1\}^n \to \mathcal{G}$,*

$$\rho(f) \geq \min(\delta/4, \eta),$$

where δ is the distance between f and the set of direct sums.

Lemma 2.3 (Iterative decoding)**.** *There exists a universal constant $\eta \in [0,1]$ such that for every function $f : \{0,1\}^n \to \mathcal{G}$ with $\rho(f) < \eta$, there exists a function $\phi : \{0,1\}^n \to \mathcal{G}$ such that:*

(i) the function ϕ is $2\rho(f)$-close to f, and

(ii) $\rho(\phi) \leq \rho(f)/2$.

Proof of Proposition 2.2. Iteratively applying Lemma 2.3 results in a sequence of functions $f = f_0, f_1, \ldots$, such that for all $t \geq 1$, (i) the distance between f_t and f_{t-1} is at most $2\rho(f_{t-1})$, and (ii) $\rho(f_t) \leq \rho(f_{t-1})/2$. The probability that the test rejects a function is a discrete quantity. So, by (ii), there must exist an integer t such that $\rho(f_t) = 0$. That is, $D_{f_t}(\pi; x, y) = 0$ for every choice of $x, y, \pi \in \{0,1\}^n$. By Fact 2.1 this means f_t is a direct sum. The distance between f and the direct sum f_t at most

$$\sum_{i=0}^{t-1} 2\rho(f_i) \leq 2\sum_{i=0}^{t-1} \rho(f)/2^i \leq 4\rho(f).$$ □

3 Analysis of Iterative Decoding

We begin with a sketch of the proof of Lemma 2.3. As mentioned in the introduction, the proof follows in the footsteps of the analysis of the BLR linearity test. We define $\phi(x)$ to be $\text{plurality}_{y,\pi}\, f(x_\pi y) + f(y_\pi x) - f(y)$. Markov's inequality allows us to bound the distance between ϕ and f by $2\rho(f)$.

To show that Test 1 rejects ϕ with probability at most $\rho(f)/2$, we first show that for all but a $o(\rho(f))$-fraction of choices of x, $\phi(x)$ is defined by a strict majority that makes up at least 6/7-th of the plurality vote (See Claim 3.4). The fraction of x's that contribute to the plurality is at least the probability of a collision, i.e., $\Pr_{y,z,\sigma,\pi}[f(x_\pi y) + f(y_\pi x) - f(y) = f(x_\sigma z) + f(z_\sigma x) - f(z)]$. Using the algebraic identity in Claim 3.2, we can express this probability as

$$\Pr_{y,z,\pi,\sigma}[D_f(\pi; x_\pi z_{\overline{\pi}}, y) - D_f(\pi; y_\pi x_{\overline{\pi}}, z) + D_f(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) = 0].$$

The analysis of the BLR test also uses an analogous algebraic identity to bound the collision probability. The difference is that the resulting expression in the BLR analysis is made up of evaluations of the BLR test at points independent of x. This allows one to argue that the plurality vote is made up of a strict majority at all points x. In our setting, the arguments of D_f in the expression above are correlated with x. However, we can view these arguments as the result of passing x through a noisy binary erasure channel. This allows for the application of the hypercontractive inequality to bound the fraction of x for which the collision probability is less than 6/7.

We then show that for all but $o(\rho(f))$ choices of $x, y, \pi \in \{0,1\}^n$ there exist $z, w, \sigma \in \{0,1\}^n$ such that, (A) the value of $D_\phi(\pi; x, y) = \phi(x) - \phi(x_\pi y) - \phi(y_\pi x) + \phi(y)$ does not change after the following substitutions, and (B) the resulting expression post substitution evaluates to zero.

$$\begin{aligned}
\phi(x) &\leftarrow f(x_\sigma z) + f(z_\sigma x) - f(z)\\
\phi(x_\pi y) &\leftarrow f((x_\pi y)_\sigma(z_\pi w)) + f((z_\pi w)_\sigma(x_\pi y)) - f(z_\pi w)\\
\phi(y_\pi x) &\leftarrow f((y_\pi x)_\sigma(w_\pi z)) + f((w_\pi z)_\sigma(y_\pi x)) - f(w_\pi z)\\
\phi(y) &\leftarrow f(y_\sigma w) + f(w_\sigma y) - f(y).
\end{aligned} \tag{2}$$

It follows that the probability that ϕ is rejected by the test is $o(\rho(f))$.

By Claim 3.4, for all but $o(\rho(f))$ choices of x, y, π the substitutions do not change the value of $D_\phi(\pi; x, y)$ with probability at least 4/7. For (B), we use the algebraic identity in Claim 3.3 to rewrite the expression after substitution as

$$D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) - D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) + D_f(\pi; z, w).$$

Again we show that the arguments of the D_f terms can be viewed as the result of passing x, y and π through independent binary erasure channels. Using the hypercontractive inequality, we conclude that for all but $o(\rho(f))$ choices of x, y and π the expression after substitution evaluates to zero for most choices of z, w and σ. By a union bound we can ensure that (A) and (B) hold simultaneously for the same z, w and σ.

The following technical lemma establishes the bounds we prove using the hypercontractivity of the binary erasure channel. The proof is presented in Section 3.1.

Let $\textsc{queries}(\pi; x, y)$ denote the vector in $(\{0,1\}^n)^4$ whose entries are the four queries that Algorithm 1 makes when π, x, y is sampled. That is, $\textsc{queries}(\pi; x, y) = (x, x_\pi y, y_\pi x, y)$.

Lemma 3.1. *Let* $\textsc{Bad} \subset (\{0,1\}^n)^4$ *be a set such that the probability that* $\textsc{queries}(\pi, x, y)$ *lands in* $\textsc{Bad}$*, when* π, x, y *are chosen independently and uniformly at random, is* ρ*.*

(i) $\mu_x(A_1) \le 21^2 \rho^{4/3}$*, where*

$$A_1 = \{x \mid \Pr_{\pi,y,z}[\textsc{queries}(\pi; x_\pi z_{\overline{\pi}}, y) \in \textsc{Bad}] \ge 1/21\}.$$

(ii) $\mu_x(A_2) \le 21^2 \rho^{4/3}$*, where*

$$A_2 = \{x \mid \Pr_{\pi,\sigma,z}[\textsc{queries}(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) \in \textsc{Bad}] \ge 1/21\}.$$

(iii) $\mu_{\pi,x,y}(A_3) \le 7^2 \rho^{2/(1+\sqrt{2/3})} \le 7^2 \rho^{1.1}$*, where*

$$A_3 = \{(\pi, x, y) \mid \Pr_{\sigma,z,w}[\textsc{queries}(\pi \oplus \sigma; x_\sigma z, y_\sigma w) \in \textsc{Bad}] \ge 1/7\}.$$

(iv) $\mu_\pi(A_4) \le 7^2 \rho^{4/3}$*, where*

$$A_4 = \{\pi \mid \Pr_{z,w}[\textsc{queries}(\pi; z, w) \in \textsc{Bad}] \ge 1/7\}.$$

We will also need the following algebraic identities.

Claim 3.2. *The following identity holds:*

$$D_f(\pi; x, y) - D_f(\sigma; x, z) = D_f(\pi; x_\pi z_{\overline{\pi}}, y) - D_f(\pi; y_\pi x_{\overline{\pi}}, z) + D_f(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}).$$

Proof of Claim 3.2. The claim follows by adding the following two identities:

$$\begin{aligned} D_f(\pi; x, y) - D_f(\pi; x, z) &= -f(x_\pi y_{\overline{\pi}}) - f(y_\pi x_{\overline{\pi}}) + f(y) + f(x_\pi z_{\overline{\pi}}) + f(z_\pi x_{\overline{\pi}}) - f(z) \\ &= f(x_\pi z_{\overline{\pi}}) + f(y) - f(x_\pi y_{\overline{\pi}}) - f(y_\pi z_{\overline{\pi}}) \\ &\quad - f(y_\pi x_{\overline{\pi}}) - f(z) + f(y_\pi z_{\overline{\pi}}) + f(z_\pi x_{\overline{\pi}}) \\ &= D_f(\pi; x_\pi z_{\overline{\pi}}, y) - D_f(\pi; y_\pi x_{\overline{\pi}}, z) \end{aligned}$$

$$\begin{aligned} D_f(\pi; x, z) - D_f(\sigma; x, z) &= -f(x_{\pi\sigma} x_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} z_{\overline{\pi\sigma}}) - f(z_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} x_{\overline{\pi\sigma}}) + f(x) + f(z) \\ &\quad + f(x_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} z_{\overline{\pi\sigma}}) + f(z_{\pi\sigma} x_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} x_{\overline{\pi\sigma}}) - f(x) - f(z) \\ &= D_f(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) \end{aligned}$$

□

To analyze the substitutions (2) we set $D_{\phi,f}(\pi; x, y) = \phi(x) - f(x_\pi y) - f(y_\pi x) + f(x)$. In particular, $D_{f,f} = D_f$.

Claim 3.3 (16-point identity)**.** *The following identity holds:*

$$\begin{aligned} &D_{\phi,f}(\sigma; x, z) - D_{\phi,f}(\sigma; x_\pi y, w_\pi z) - D_{\phi,f}(\sigma; y_\pi x, z_\pi w) + D_{\phi,f}(\sigma; y, w) \\ &\qquad = D_\phi(\pi; x, y) - D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) - D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) + D_f(\pi; z, w). \end{aligned}$$

Proof of Claim 3.3. We write $xyzw$ to denote the string $x_{\pi\sigma} y_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} w_{\overline{\pi\sigma}}$. With this notation,

$$\begin{array}{lllll}
+D_{\phi,f}(\sigma; x, z) & = +\phi(xxxx) & -f(xzxz) & -f(zxzx) & +f(zzzz) \\
-D_{\phi,f}(\sigma; x_\pi y, w_\pi z) & = -\phi(xxyy) & +f(xwyz) & +f(wxzy) & -f(wwzz) \\
-D_{\phi,f}(\sigma; y_\pi x, z_\pi w) & = -\phi(yyxx) & +f(yzxw) & +f(zywx) & -f(zzww) \\
+D_{\phi,f}(\sigma; y, w) & = +\phi(yyyy) & -f(ywyw) & -f(wywy) & +f(wwww) \\
 & \quad \| & \quad \| & \quad \| & \quad \| \\
 & +D_\phi(\pi; x, y) & -D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) & -D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) & +D_f(\pi; z, w)
\end{array}$$

The identity states that the column sums and the row sums add up. □

Proof of Lemma 2.3. Let ϕ be a function defined as $\phi(x) = \text{plurality}_{y,\pi} f(x_\pi y) + f(y_\pi x) - f(y)$. That is, $\phi(x)$ is the most frequent value of $f(x_\pi y) + f(y_\pi x) - f(y)$, where $y, \pi \in \{0,1\}^n$. Ties are broken arbitrarily. We show that ϕ satisfies the hypothesis of the lemma.

(i) ϕ is $2\rho(f)$-close to f: For $x \in \{0,1\}^n$, let $\rho_x \doteq \Pr_{y,\pi}[f(x) \neq f(x_\pi y) + f(y_\pi x) - f(y)]$. Note that $\mathbb{E}_x[\rho_x] = \rho(f)$, and that if $\rho_x < 1/2$ then $f(x) = \phi(x)$. Thus, by Markov's inequality,

$$\Pr_x[f(x) \neq \phi(x)] \leq \Pr_x[\rho_x \geq 1/2] \leq 2\rho(f).$$

(ii) $\rho(\phi) \leq \rho(f)/2$: We begin by showing that with probability $\rho(f)/12$ over the choice of x, the plurality that defines $\phi(x)$ is a majority made up of 6/7-th of the votes. Then $\Pr_{\pi,y}[D_{\phi,f}(\pi; x, y) = 0]$ is the fraction of votes that constitute the plurality defining $\phi(x)$. Let

$$\textsc{Weak-Maj} = \big\{x \;\big|\; \Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) \neq 0] \geq 1/7\big\}.$$

Claim 3.4 (Strong Majority)**.** $\mu_x(\textsc{Weak-Maj}) \leq \rho(f)/12$.

Proof. The fraction of votes that contribute to the plurality $\Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) = 0]$ is an upper bound on the collision probability $\Pr_{y,\pi,z,\sigma \in \{0,1\}^n}[D_{\phi,f}(\pi; x, y) = D_{\phi,f}(\sigma; x, z)]$. This is because

$$\begin{aligned}
\Pr_{y,\pi,z,\sigma}[D_{\phi,f}(\pi; x, y) = D_{\phi,f}(\sigma; x, z)] &= \sum_{\gamma \in \mathcal{G}} \Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) = \gamma]^2 \\
&\leq \max_{\gamma \in \mathcal{G}} \Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) = \gamma] \\
&= \Pr_{y,\pi}[D_{\phi,f}(\pi; x, y) = 0].
\end{aligned}$$

The final equality holds because $\phi(x) = \arg\max_{\beta \in \mathcal{G}} \Pr_{y,\pi}[\beta - f(x_\pi y) - f(y_\pi x) + f(y) = 0]$. We showed that

$$\mu_x(\textsc{Weak-Maj}) \leq \mu_x\{x \mid \Pr_{y,\pi,z,\sigma}[D_{\phi,f}(\pi; x, y) \neq D_{\phi,f}(\sigma; x, z)] \geq 1/7\}.$$

We now use Lemma 3.1 to bound the right hand side. Let $\textsc{Bad}_f \subset (\{0,1\}^n)^4$ be the set of queries on which D_f fails, namely

$$\textsc{Bad}_f = \{\textsc{queries}(\pi; x, y) \mid D_f(\pi; x, y) \neq 0\}.$$

This is a set of measure $\mu_{\pi,x,y}(\textsc{Bad}) = \rho(f)$. Since $D_{\phi,f}(\pi; x, y) - D_{\phi,f}(\sigma; x, z) = D_f(\pi; x, y) - D_f(\sigma; x, z)$, by the algebraic identity in Claim 3.2 and a union bound, we have

$$\begin{aligned}
\mu_x(\textsc{Weak-Maj}) &\leq \mu_x\{x \mid \Pr_{\pi,\sigma,y,z}[D_{\phi,f}(\pi; x, y) - D_{\phi,f}(\sigma; x, z) \neq 0] \geq 1/7\} \\
&\leq \mu_x\{x \mid \Pr_{\pi,y,z}[D_f(\pi; x_\pi z_{\overline{\pi}}, y) \neq 0] \geq 1/21\} \\
&\quad\;+\; \mu_x\{x \mid \Pr_{\pi,y,z}[D_f(\pi; y_\pi x_{\overline{\pi}}, z) \neq 0] \geq 1/21\} \\
&\quad\;+\; \mu_x\{x \mid \Pr_{\pi,\sigma,z}[D_f(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) \neq 0] \geq 1/21\} \\
&= \mu_x(A_1) + \mu_x(A_1) + \mu_x(A_2),
\end{aligned}$$

where A_1 and A_2 are the sets

$$\begin{aligned}
A_1 &= \{x \mid \Pr_{\pi,y,z}[\textsc{queries}(\pi; x_\pi z_{\overline{\pi}}, y) \in \textsc{Bad}_f] \geq 1/21\}, \\
A_2 &= \{x \mid \Pr_{\pi,\sigma,z}[\textsc{queries}(\pi \oplus \sigma; x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) \in \textsc{Bad}_f] \geq 1/21\}.
\end{aligned}$$

By Lemma 3.1 we get that $\mu_x(\textsc{Weak-Maj}) \leq \rho(f)/12$, for small enough η. □

We are now ready to prove that $\rho(\phi) = \Pr_{\pi,x,y}[D_\phi(\pi; x, y) \neq 0] \leq \rho(f)/2$. In order to do so we define a set $\textsc{Bad}_\phi$ of triples (π, x, y) such that $\mu_{\pi,x,y}(\textsc{Bad}_\phi) \leq \rho(f)/2$, and if $(\pi, x, y) \notin \textsc{Bad}_\phi$ then $D_\phi(\pi, x, y) = 0$.

Let A_3 and A_4 denote the sets

$$\begin{aligned}
A_3 &= \{(\pi, x, y) \mid \Pr_{\sigma,z,w}[\textsc{queries}(\pi \oplus \sigma; x_\sigma z, y_\sigma w) \in \textsc{Bad}_f] \geq 1/7\}, \\
A_4 &= \{\pi \mid \Pr_{z,w}[\textsc{queries}(\pi; z, w) \in \textsc{Bad}_f] \geq 1/7\}.
\end{aligned}$$

Let $\textsc{Bad}_\phi$ be the set

$$\{(\pi, x, y) \mid (\text{One of } x, y, x_\pi y, y_\pi x \text{ lies in } \textsc{Weak-Maj}) \text{ or } ((\pi, x, y) \in A_3) \text{ or } (\pi \in A_4)\}\,.$$

By Lemma 3.1, $\mu_{\pi,x,y}(A_3) \leq \rho(f)/12$, and $\mu_\pi(A_4) \leq \rho(f)/12$, for small enough η. As $x, y, x_\pi y, y_\pi x$ are all random, by a union bound we have

$$\mu_{\pi,x,y}(\textsc{Bad}_\phi) \leq 4\mu_x(\textsc{Weak-Maj}) + \mu_{\pi,x,y}(A_3) + \mu_\pi(A_4) \leq \rho(f)/2.$$

All that remains to show is that if $(\pi, x, y) \notin \textsc{Bad}_\phi$, $D_\phi(\pi; x, y) = 0$. On rearranging the terms in the algebraic identity of Claim 3.3, we get

$$\begin{aligned}
D_\phi(\pi; x, y) = \;& D_{\phi,f}(\sigma; x, z) - D_{\phi,f}(\sigma; x_\pi y, w_\pi z) - D_{\phi,f}(\sigma; y_\pi x; z_\pi w) + D_{\phi,f}(\sigma; y, w) \\
& + D_f(\pi \oplus \sigma; x_\sigma z, y_\sigma w) + D_f(\pi \oplus \overline{\sigma}; x_{\overline{\sigma}} z, y_{\overline{\sigma}} w) - D_f(\pi; z, w).
\end{aligned} \tag{3}$$

Fix a triple $(\pi, x, y) \notin \textsc{Bad}_\phi$. We show that $D_\phi(\pi, x, y) = 0$, by showing that there exists a choice of σ, z and w for which the right hand side of Equation (3) evaluates to zero. By Equation (3) and a union bound,

$$\begin{aligned}\Pr_{\sigma,z,w}[D_\phi(\pi;x,y)\neq 0] &= \Pr_{\sigma,z,w}\left[\begin{array}{l} \quad D_{\phi,f}(\sigma;x,z)\neq 0 \\ \text{or } D_{\phi,f}(\sigma;x_\pi y,w_\pi z)\neq 0 \\ \text{or } D_{\phi,f}(\sigma;y_\pi x;z_\pi w)\neq 0 \\ \text{or } D_{\phi,f}(\sigma;y,w)\neq 0 \\ \text{or } D_f(\pi\oplus\sigma;x_\sigma z,y_\sigma w)\neq 0 \\ \text{or } D_f(\pi\oplus\overline{\sigma};x_{\overline{\sigma}} z,y_{\overline{\sigma}} w)\neq 0 \\ \text{or } D_f(\pi;z,w)\neq 0 \end{array}\right] \\ &< 4/7+\Pr_{\sigma,z,w}\left[\begin{array}{l} \quad \textsc{queries}(\pi\oplus\sigma;x_\sigma z,y_\sigma w)\in\textsc{Bad}_f \\ \text{or } \textsc{queries}(\pi\oplus\overline{\sigma};x_{\overline{\sigma}} z,y_{\overline{\sigma}} w)\in\textsc{Bad}_f \\ \text{or } \textsc{queries}(\pi;z,w)\in\textsc{Bad}_f \end{array}\right] \\ &< 4/7+3/7=1.\end{aligned}$$

The first inequality holds because $x, x_\pi y, y_\pi x, y \notin$ Weak-Maj, and the second inequality holds because $(\pi, x, y) \notin A_3$ and $\pi \notin A_4$. Since the probability $\Pr_{\sigma,z,w}[D_\phi(\pi;x,y)\neq 0]$ is either 0 or 1, it must be that $D_\phi(\pi;x,y)=0$. Therefore,

$$\rho(\phi)=\Pr_{\pi,x,y}[D_\phi(\pi;x,y)\neq 0]\le \mu_{\pi,x,y}(\textsc{Bad}_\phi)\le \rho(f)/2. \qquad \square$$

3.1 Proof of Lemma 3.1

We begin with some preliminaries on discrete channels and hypercontractivity. For a motivating discussion on hypercontractivity and a proof of Fact 3.8 below see Chapter 9 of [O'D14].

Definition 3.5 (Discrete channels)**.** A *discrete channel* is a triple $(\mathcal{U}, P, \mathcal{V})$, where $\mathcal{U}$ and $\mathcal{V}$ are finite sets representing the *input alphabet* and *output alphabet*, and P is a $\mathcal{U}\times\mathcal{V}$ probability *transition matrix* that describes the distribution of the output conditioned on the input. The *composition* of two channels $(\mathcal{U}, P_1, \mathcal{V})$ and $(\mathcal{V}, P_2, \mathcal{W})$ is the channel $(\mathcal{U}, P_1\cdot P_2, \mathcal{W})$, where $\cdot$ is matrix multiplication.

The binary erasure channel will play an important role in the proof of Lemma 3.1.

Definition 3.6 (Binary Erasure Channel)**.** The *binary erasure channel* $BEC(e)$ with erasure probability e has input alphabet $\{0,1\}$ and output alphabet $\{0,1,\perp\}$, and probability transition matrix $P(x|x)=1-e, P(\perp|x)=e$.

For a real valued random variable U and $p\ge 1$, we denote the p-norm of U by $\|U\|_p \doteq \mathbb{E}_U[|U|^p]^{1/p}$.

Definition 3.7 (Hypercontractivity)**.** For $1\le q\le p$, A pair of random variables (U,V) is (p,q)-hypercontractive if for every pair of real valued functions f, g,

$$\mathbb{E}[f(U)g(V)]\le \|f(U)\|_{p'}\|\|g(V)\|_q,$$

where $p'=p/(p-1)$ is the Hölder conjugate of p.

Fact 3.8 (Tensorisation [Bon70])**.** *If (U_1,V_1) and (U_2,V_2) are independent random variables that are (p,q)-hypercontractive, then $((U_1,U_2),(V_1,V_2))$ is (p,q) hypercontractive.*

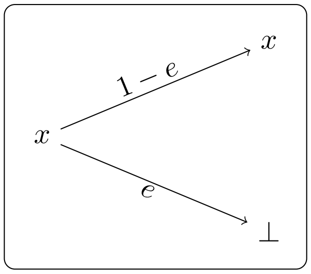

Figure 1: The binary erasure channel $BEC(e)$.

Theorem 3.9 (Hypercontractivity of $BEC(e)$ [NW16])**.** *Let U be distributed uniformly over $\{0,1\}$ and let $V \in \{0,1,\perp\}$ denote the output of $BEC(e)$ on input U. Then (U,V) is (p,q)-hypercontractive for all $1 \le q \le p$ such that*

$$\frac{q-1}{p-1} \ge 1-e.$$

Fact 3.10 (Composition)**.** *Let $(\mathcal{U}, P_1, \mathcal{V})$ and $(\mathcal{V}, P_2, \mathcal{W})$ be two channels. Let U be a random variable over $\mathcal{U}$. Let V be the random variable that represents the output of the first channel on input U, and W the random variable that represents the output of the second channel on input V. If (U,V) is (p,q)-hypercontractive then so is (U,W).*

Proof of Fact 3.10. Let $f : \mathcal{U} \to \mathbb{R}$ and $g : \mathcal{W} \to \mathbb{R}$ be arbitrary functions. Since $U \to V \to W$ is a markov chain, we have

$$\mathbb{E}_{U,W}[f(U)g(W)] = \mathbb{E}_{U,V}[f(U)E_W[g(W) \mid V]] \le \|f(U)\|_{p'} \|\mathbb{E}_W[g(W) \mid V]\|_q,$$

where the inequality holds because (U,V) is (p,q) hypercontractive.

Now, by Jensen's inequality,

$$\mathbb{E}_V[\mathbb{E}_W[g(W) \mid V]^q]^{1/q} \le \mathbb{E}_V[\mathbb{E}_W[g(W)^q \mid V]]^{1/q} = \mathbb{E}_W[g(W)^q]^{1/q} = \|g(W)\|_q$$

Therefore, (U,W) is (p,q) hypercontractive. □

The following claim captures the small-set expansion interpretation of hypercontractivity [AG76] in the form used in the proof of Lemma 3.1.

Claim 3.11. *Let U,V be random variables that take values in $\mathcal{U}$ and $\mathcal{V}$ respectively. Let $B \subset \mathcal{V}$ be a set such that $\Pr[V \in B] = \rho$. Let $A \subset \mathcal{U}$ denote the set $\{u \mid \Pr[V \in B \mid U = u] \ge \theta\}$. If (U,V) is (p,q) hypercontractive, then $\Pr[U \in A] \le \rho^{p/q}/\theta^p$.*

Proof. Let 1_A and 1_B denote the indicator functions of the sets A and B. Since (U,V) are (p,q) hypercontractive,

$$\theta \cdot \Pr[U \in A] \le \Pr[V \in B \mid U \in A] \Pr[U \in A] = \mathbb{E}[1_A(U) 1_B(V)] \le \|1_A(U)\|_{p'} \|1_B(V)\|_q,$$

where $p' = p/(p-1)$. Note that $\|1_B(V)\| = \rho^{1/q}$, and $\|1_A(U)\|_{p'} = \Pr[U \in A]^{1/p'}$. Therefore, $\Pr[U \in A]^{1/p} \le \rho^{1/q}/\theta$, that is, $\Pr[U \in A] \le \rho^{p/q}/\theta^p$. □

Proof of Lemma 3.1. We need to bound the probabilities of four sets of the form

$$\{u \in \Sigma^n | \Pr[\textsc{queries}(\psi(u)) \in \textsc{bad} | U = u]\} \geq \theta,$$

where ψ is some (randomized) function. All bounds of the form $\theta^{-2}\rho^{2/q}$ will follow from Claim 3.11 by showing that the channel $U \to \textsc{queries}(\psi(u))$ is $(2,q)$-hypercontractive for a suitable choice of q ($q = 3/2$ for parts (i), (ii), (iv) and $q = 1 + \sqrt{2/3}$ for part (iii)).

The *channel* $(\Sigma^n, P_n, \{0,1\}^{n\times 4})$ that maps $u \in \Sigma^n$ to $\textsc{queries}(\psi(u)) \in \{0,1\}^{n\times 4}$ acts independently on the symbols $u_1, \ldots, u_n$. In all cases, the i-th bits of the four queries (q_1, q_2, q_3, q_4) are obtained by applying the one-dimensional channel P_1 to u_i. Therefore, P_n tensorizes as $P_n = P_1^{\otimes n}$. By Fact 3.8, it is sufficient to show that the channel P_1 is hypercontractive. We may and will therefore assume, without loss of generality, that $n = 1$.

We now demonstrate how each of the four channels of interest can be decomposed into a binary erasure channel with constant erasure probability ($e = 1/2$ in parts (i), (ii), (iv) and $e = 1 - \sqrt{2/3}$ in part (iii)) and some other fixed channel. The Lemma then follows from Fact 3.10 and Theorem 3.9 with $q = 2 - e$.

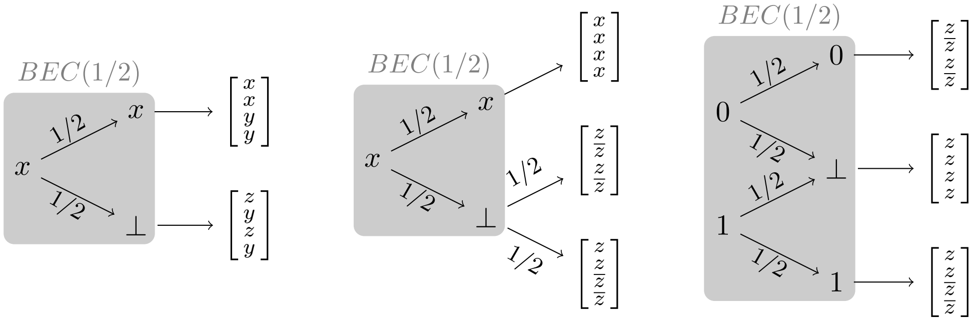
$BEC(1/2)$
x
$1/2$
x
$1/2$
$\perp$
$\begin{bmatrix} x \\ x \\ y \\ y \end{bmatrix}$
$\begin{bmatrix} z \\ y \\ z \\ y \end{bmatrix}$
$BEC(1/2)$
x
$1/2$
x
$1/2$
$\perp$
$1/2$
$1/2$
$\begin{bmatrix} x \\ x \\ x \\ x \end{bmatrix}$
$\begin{bmatrix} z \\ \overline{z} \\ z \\ \overline{z} \end{bmatrix}$
$\begin{bmatrix} z \\ z \\ \overline{z} \\ \overline{z} \end{bmatrix}$
$BEC(1/2)$
0
$1/2$
0
$1/2$
$\perp$
1
$1/2$
$1/2$
1
$\begin{bmatrix} z \\ \overline{z} \\ z \\ \overline{z} \end{bmatrix}$
$\begin{bmatrix} z \\ z \\ z \\ z \end{bmatrix}$
$\begin{bmatrix} z \\ z \\ \overline{z} \\ \overline{z} \end{bmatrix}$

Figure 2: Channels (i) $x \to \textsc{queries}(\pi, x_\pi z_{\overline{\pi}}, y)$; (ii) $x \to \textsc{queries}(\pi \oplus \sigma, x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}})$; (iv) $\pi \to \textsc{queries}(\pi, z, w)$. y and z are random bits.

(i) The channel $x \to \textsc{queries}(\pi; x_\pi z_{\overline{\pi}}, y) = (x_\pi z_{\overline{\pi}}, x_\pi y_{\overline{\pi}}, y_\pi z_{\overline{\pi}}, y)$ from $\Sigma = \{0,1\}$ to $\{0,1\}^4$ can be decomposed in the following way: On input x the channel samples a random bit π and outputs $xxyy$ if $\pi = 1$, and $zyzy$ if $\pi = 0$ for random y and z. This channel can be alternatively described as $BEC(1/2)$ composed with a second channel that outputs $xxyy$ if there is no erasure and the independent symbol $zyzy$ otherwise. See Figure 2 (i).

(ii) The channel from $\Sigma = \{0,1\}$ to $\{0,1\}^4$ is of the form

$$x \to \textsc{queries}(\pi \oplus \sigma, x_\sigma z_{\overline{\sigma}}, z_\sigma x_{\overline{\sigma}}) = \begin{pmatrix} x_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} z_{\overline{\pi\sigma}} \\ z_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} x_{\overline{\pi\sigma}} \\ x_{\pi\sigma} x_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} z_{\overline{\pi\sigma}} \\ z_{\pi\sigma} x_{\pi\overline{\sigma}} z_{\overline{\pi}\sigma} x_{\overline{\pi\sigma}} \end{pmatrix} = \begin{cases} xzxz, & \text{if } \pi\sigma = 1, \\ zzxx, & \text{if } \pi\overline{\sigma} = 1, \\ xxzz, & \text{if } \overline{\pi}\sigma = 1, \\ zxzx, & \text{if } \overline{\pi\sigma} = 1, \end{cases}$$

where π, σ, z are random bits. We can alternatively describe it like this: If $z = x$, then output $xxxx$. If $z \neq x$ and $\pi \oplus \sigma = 1$, then output $zz\overline{zz}$. If $z \neq x$ and $\pi \oplus \sigma = 0$, then output $zzzz$.

This channel can be factored through $BEC(1/2)$ as in Figure 2 (ii). If there is no erasure, the second channel outputs $xxxx$. If there is an erasure, then the second channel outputs $zz\overline{z}\overline{z}$ with probability $1/2$ and $z\overline{z}z\overline{z}$ with probability $1/2$.

(iii) The channel from $\Sigma = \{0,1\}^3$ to $\{0,1\}^4$ is of the form

$$\begin{pmatrix}\pi\\ x\\ y\end{pmatrix} \to \textsc{queries}(\pi \oplus \sigma; x_\sigma z, y_\sigma w) = \begin{pmatrix} x_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} z_{\overline{\pi\sigma}} \\ y_{\pi\sigma} z_{\pi\overline{\sigma}} x_{\overline{\pi}\sigma} w_{\overline{\pi\sigma}} \\ x_{\pi\sigma} w_{\pi\overline{\sigma}} y_{\overline{\pi}\sigma} z_{\overline{\pi\sigma}} \\ y_{\pi\sigma} w_{\pi\overline{\sigma}} y_{\overline{\pi}\sigma} w_{\overline{\pi\sigma}} \end{pmatrix} = \begin{cases} xyxy, & \text{if } \pi\sigma = 1, \\ zzww, & \text{if } \pi\overline{\sigma} = 1, \\ xxyy, & \text{if } \overline{\pi}\sigma = 1, \\ zwzw, & \text{if } \overline{\pi\sigma} = 1. \end{cases}$$

Consider the composition of the following two channels. The first channel views the symbol $\pi x y$ as three bits and independently applies $BEC(1/4)$ to π and $BEC(1-\sqrt{2/3})$ to x and y. The second channel is described in Figure 3.

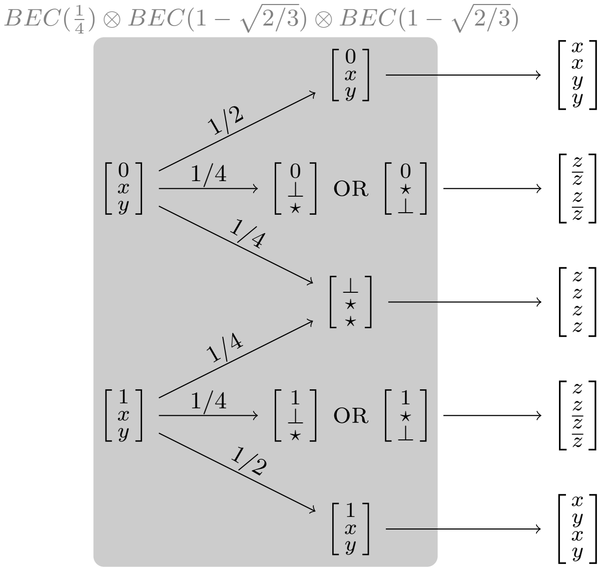
$BEC(\frac{1}{4}) \otimes BEC(1-\sqrt{2/3}) \otimes BEC(1-\sqrt{2/3})$
$[0, x, y]$
1/2
1/4
1/4
$[0, x, y]$
$[0, \perp, \star]$ OR $[0, \star, \perp]$
$[\perp, \star, \star]$
$[1, x, y]$
1/4
1/4
1/2
$[1, \perp, \star]$ OR $[1, \star, \perp]$
$[1, x, y]$
$[x, x, y, y]$
$[z, \overline{z}, z, \overline{z}]$
$[z, z, z, z]$
$[z, z, \overline{z}, \overline{z}]$
$[x, y, x, y]$

Figure 3: (iii) Channel $(\pi, x, y) \to \textsc{queries}(\pi \oplus \sigma, x_\sigma z, y_\sigma w)$. A $\star$ represents any of $\{0, 1, \perp\}$ and z is a random bit.

- If π is erased, the second channel outputs $zzzz$, for a uniform bit z. This corresponds to the event $z = w$ and $\sigma = 0$.
- If π is not erased but one of x, y is erased, the second channel samples a uniform bit $z \in \{0,1\}$ and outputs $z\overline{z}z\overline{z}$ if $\pi = 0$ and $zz\overline{z}\overline{z}$ if $\pi = 1$. This corresponds to the event $z \neq w$ and $\sigma = 0$.
- If there are no erasures, then the second channel outputs $xyxy$ if $\pi = 1$, and $xxyy$ if $\pi = 0$. This corresponds to the event $\sigma = 1$.

The first channel is $BEC(\frac{1}{4}) \otimes BEC(1-\sqrt{2/3}) \otimes BEC(1-\sqrt{2/3})$. Since $1 - \sqrt{2/3} \leq 1/4$, by Fact 3.8 it inherits the hypercontractivity parameters of $BEC(1-\sqrt{2/3})$.

(iv) The channel $\pi \to \textsc{queries}(\pi, z, w)$ from $\Sigma = \{0,1\}$ to $\{0,1\}^4$ outputs $zzww$ if $\pi = 1$ and $zwzw$ if $\pi = 0$ for random bits z and w. Alternatively, the channel can be described as a uniform choice between $zzzz$ and $zz\overline{z}\overline{z}$ when $\pi = 1$ and a uniform choice between $zzzz$ and $z\overline{z}z\overline{z}$ when $\pi = 0$. This can be modeled as the composition of $BEC(1/2)$ and a second channel that outputs $zzzz$ if there is an erasure, and either $zz\overline{z}\overline{z}$ or $z\overline{z}z\overline{z}$ depending on the value of π_i otherwise. See Figure 2 (iv). □

4 From the Hypercube to Arbitrary Product Domains

In this section we prove Theorem 1.1. Recall that for a function $f : \mathcal{D}_1 \times \ldots \times \mathcal{D}_n \to \mathcal{G}$, $\delta(f)$ is the distance between f and the set of direct sums, and $\rho(f)$ is equal to $\Pr_{x,y,S,T}[D_f(S,T;x,y) \neq 0]$, where the probability is over random x, y and a uniformly random partition (S,T) of $[n]$. Theorem 1.1 follows directly from Proposition 2.2 and the following Proposition.

Proposition 4.1. *If $\delta(f) = O(\rho(f))$ for every $f\colon \{0,1\}^n \to \mathcal{G}$ then $\delta(f) = O(\rho(f))$ for every $f\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$.*

The proof of Proposition 4.1 uses the following equivalent description of direct sums over $\{0,1\}^n$.

Claim 4.2. *$f\colon \{0,1\}^n \to \mathcal{G}$ is a direct sum if and only if $f(u_1,\ldots,u_n) = u_1g_1 + \cdots + u_ng_n + g_0$ for some $g_0, g_1, \ldots, g_n \in \mathcal{G}$, where $0g = 0 \in \mathcal{G}$ and $1g = g$.*

Proof. The if direction is immediate. For the only if direction, if $f(u) = f_1(u_1) + \cdots + f_n(u_n)$, write $f_i(u_i) = u_i(f_i(1) - f_i(0)) + f_i(0)$ to obtain the desired representation with $g_i = f_i(1) - f_i(0)$ and $g_0 = f_1(0) + \cdots + f_n(0)$. □

To prove Proposition 4.1 we first analyze the soundness of the following alternative test T_f. The test T_f runs the test D_g on the Boolean function g defined as the restriction of f to the "slice" $\{x_1, y_1\} \times \cdots \times \{x_n, y_n\}$, where x_i, y_i are random samples from $\mathcal{D}_i$.

```
Algorithm 2: Direct sum test $T_f$
  Oracle : $f\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$
  Pick $x, y$ at random from $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$.
  Let $g\colon \{0,1\}^n \to \mathcal{G}$ be the function $g(\tau; x, y) = f(x_\tau y_{\overline{\tau}})$.
  Accept if $D_{g(\cdot;x,y)} = 0$ on random inputs.
```

The rejection probability of T has the following useful alternative characterization. For arbitrary disjoint subsets U, V of $[n]$, let $D_f(U,V;x,y) = f(x) - f(y_Ux) - f(y_Vx) + f(y_{(U\cup V)}x)$.

Fact 4.3. *Let U, V be disjoint random subsets of $[n]$ sampled by independently assigning each index $i \in [n]$ to U with probability $1/4$, to V with probability $1/4$ and to neither with probability $1/2$. Then the probability that the test T_f rejects is equal to $\Pr_{U,V,x,y}[D_f(U,V;x,y) \neq 0]$.*

If T_f accepts with probability at least $1-\epsilon$, then by averaging there exists a y such that $D_{g(\cdot;x,y)}$ accepts with probability at least $1-\epsilon$. By the soundness of the test D for boolean functions (Proposition 2.2), we conclude that for a random choice of x, f restricted to the slice defined by x and y is $O(\epsilon)$-close to a direct sum. In particular, there exist functions $g_i : \mathcal{D}_i \to \mathcal{G}$ such that $f(x_\tau y_{\overline{\tau}})$ is $O(\epsilon)$-close to $\tau_1 g_1(x) + \cdots + \tau_n g_n(x) + g_0(x)$. However, f could be close to different direct sums on different slices. Intuitively, if f is close to a direct sum, then $f(x_\tau y_{\overline{\tau}})$ should not depend on $x_{\overline{\tau}}$. In Claim 4.5, we use a variant of the Dinur-Steurer direct product test [DS14] to show that this is indeed the case. We show that the restriction $f(x_\tau y_{\overline{\tau}})$ is $O(\epsilon)$-close to a function of the form $\sum_{i=1}^n \tau_i \tilde{g}_i(x_i) + \tilde{g}_0$. In Claim 4.7, we show that this implies f is $O(\epsilon)$-close to a direct sum on the entire domain. In other words, T_f is sound. To show soundness of the test D_f on general domains, we upper bound the rejection probability of T_f with that of D_f (Claim 4.10).

We use the following variant of the Dinur-Steurer direct product test [DS14] (see also Section 2.2 in [DG19]): Given oracle access for a multivariate function $G\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}^n$, choose two random strings x and y and a random subset S of $[n]$ by including each element independently with probability $1/3$ and accept if $G(x)$ and $G(y_S x)$ match on all outputs outside S.

Let (S, T) be random disjoint subsets of $[n]$ where each element is independently assigned to S with probability $1/3$ and to T with probability $1/2$.

Claim 4.4. *Let $G(x) = (g_1(x), \dots, g_n(x))$ and g_0 be any function. If δ is the probability that $G(x)$ and $G(z_S x)$ differ on some output outside S, then $\sum_{i\in T} g_i(x) + g_0(x)$ differs from $\sum_{i\in T} g_i(z_S x) + g_0(z_S x)$ with probability at least $\delta/4$.*

Proof. Assume $G(x)$ and $G(z_S x)$ differ on some output $t \notin S$, namely $g_t(x) \neq g_t(z_S x)$. After fixing membership in T for all elements except t, the sum $\sum_{i\in T\cup\{0\}} g_i(x) - g_i(z_S x)$ is undetermined as it can take one of the two values h or $h + g_t(x) - g_t(z_S x)$ for some $h \in \mathcal{G}$, one of which must be nonzero. This nonzero value is taken with probability at least $\min\{\Pr[t \in T | t \notin S], \Pr[t \notin T | t \notin S]\} = 1/4$. □

Claim 4.5. *If for a fixed y, $\Pr_{\tau,x}[g(\tau; x, y) \neq \tau_1 g_1(x) + \cdots + \tau_n g_n(x) + g_0(x)] \leq \epsilon$ then there exist $\tilde{g}_i : \mathcal{D}_i \to \mathcal{G}$, for $i \in [n]$, and $\tilde{g}_0 \in \mathcal{G}$, such that $\Pr_{\tau,x}[f(x_\tau y_{\overline{\tau}}) \neq \sum_{i=1}^n \tau_i \tilde{g}_i(x_i) + \tilde{g}_0] = O(\epsilon)$.*

For random variables A and B we use the notation $A \approx_\epsilon B$ as a shorthand for $\Pr[A \neq B] \leq \epsilon$. In this notation the triangle inequality reads as

$$A \approx_\epsilon B, A \approx_{\epsilon'} B' \;\longrightarrow\; B \approx_{\epsilon+\epsilon'} B'.$$

Proof. We can write the assumption as

$$f(x_\tau y_{\overline{\tau}}) \approx_\epsilon \tau_1 g_1(x) + \cdots + \tau_n g_n(x) + g_0(x). \tag{4}$$

Let $T = \{i \colon \tau_i = 1\}$ and S be a random subset of $\overline{T}$ in which each entry is included independently with probability $2/3$. Then (S, T) are jointly distributed as in Claim 4.4. For z independent of x,

$$f((z_S x)_\tau y_{\overline{\tau}}) \approx_\epsilon \tau_1 g_1(z_S x) + \cdots + \tau_n g_n(z_S x) + g_0(z_S x).$$

As S and T are disjoint, $(z_S x)_\tau y_{\overline{\tau}} = x_\tau y_{\overline{\tau}}$, and by the triangle inequality, the right-hand sides of the two expressions are 2ϵ close:

$$\sum_{i\in T} g_i(x) + g_0(x) \approx_{2\epsilon} \sum_{i\in T} g_i(z_S x) + g_0(z_S x). \tag{5}$$

By Claim 4.4, $G(x)$ and $G(z_S x)$ differ on some output outside S with probability at most 6ϵ. By the soundness of the Dinur-Steurer test,

$$(g_1(x), \dots, g_n(x)) \approx_{O(\epsilon)} (\tilde{g}_1(x_1), \dots, \tilde{g}_n(x_n)) \tag{6}$$

for some functions $\tilde{g}_1, \dots, \tilde{g}_n$. By the triangle inequality applied to (5) and (6),

$$\sum_{i \in T} \tilde{g}_i(x_i) + g_0(x) \approx_{O(\epsilon)} \sum_{i \in T} \tilde{g}_i((z_S x)_i) + g_0(z_S x).$$

Since $(z_S x)_i$ only differs from x_i when $i \in S$, and S and T are disjoint, after rearranging terms we obtain

$$g_0(x) \approx_{O(\epsilon)} g_0(z_S x). \tag{7}$$

Let (S, S', S'') be a partition of $[n]$ where each element is independently and randomly assigned to one of the three sets. Since the pairs $(x, z_S x)$, $(z_S x, z_{S \cup S'} x)$, and $(z_{S \cup S'} x, z)$ are identically distributed, from (7) we also have

$$g_0(z_S x) \approx_{O(\epsilon)} g_0(z_{S \cup S'} x)$$
$$g_0(z_{S \cup S'} x) \approx_{O(\epsilon)} g_0(z).$$

By the triangle inequality

$$g_0(x) \approx_{O(\epsilon)} g_0(z).$$

Fixing z we obtain that $g_0(x)$ is $O(\epsilon)$-close to $\tilde{g_0} = g_0(z)$. Plugging this and (6) into (4) we obtain the desired approximation of $f(x_\tau y_{\overline{\tau}})$. □

Claim 4.6. *Let (U, V, U', V') be a partition of $[n]$. Let $S = U \cup U'$ (and $\overline{S} = V \cup V'$). Then:*

$$D_f(S, \overline{S}; x, y) = D_f(U, V; x, y) + D_f(U', V; y, x) + D_f(U, V'; y, x) + D_f(U', V'; x, y)$$

Proof. By Claim 5.2 below and the fact that $D_f(S, \overline{S}; x, y) = D_f(\overline{S}, S; x, y)$

$$D_f(S, \overline{S}; x, y) = D_f(U, \overline{S}; x, y) + D_f(U', \overline{S}; y, x)$$

Applying Claim 5.2 again to $D_f(U, \overline{S}; x, y)$ and $D_f(U', \overline{S}; x, y)$, we get the required identity. □

Let $\rho_T(f)$ denote the probability that the test T_f rejects f.

Claim 4.7 (Soundness of T_f). *If $\delta(f) = O(\rho(f))$ for every $f : \{0,1\}^n \to \mathcal{G}$ then $\delta(f) = O(\rho_T(f))$ for every $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$.*

To prove this claim, we argue that there exists a point y that satisfies the hypothesis of Claim 4.5 and

$$\Pr_{\tau, x}[f(x) \neq f(x_\tau y_{\overline{\tau}}) + f(x_{\overline{\tau}} y_\tau) - f(y)] \leq 12 \rho_T(f).$$

Using the conclusion of Claim 4.5 we get that

$$f(x) \approx_{O(\rho_T(f))} \sum_{i=1}^{n} \tilde{g}_i(x_i) + 2\tilde{g}_0 - f(y).$$

Proof of Claim 4.7. Assume $\delta(f) = O(\rho(f))$ for every $f : \{0,1\}^n \to \mathcal{G}$. Let $\epsilon = \rho_T(f)$. So,

$$D_{g(\cdot; x, y)}(S, \overline{S}; u, v) \approx_\epsilon 0.$$

By Markov's inequality, a 2/3-rd fraction of y's satisfy

$$D_{g(\cdot;x,y)}(S,\overline{S};u,v) \approx_{3\epsilon} 0, \tag{8}$$

Let $U, V \subseteq [n]$ be disjoint random sets that are obtained by adding each element of $[n]$ independently to U with probability 1/4, V with probability 1/4, and neither with probability 1/2. By Fact 4.3, the probability that the test T_f rejects f is equal to $\Pr_{U,V,x,y}[D_f(U,V;x,y) \neq 0]$. Thus by Claim 4.6 and a union bound, $\Pr_{T,x,y}[D_f(T,\overline{T};x,y) \neq 0] \leq 4 \cdot \epsilon$, where the probability is over a uniformly random bipartition $\tau = (T,\overline{T})$ of $[n]$. Again by Markov's inequality, at least 2/3 of the y satisfy

$$D_f(\tau;x,y) \approx_{12\epsilon} 0. \tag{9}$$

Fix a y that satisfies both (8) and (9). From (8) and the soundness of D_f over the Boolean domain,

$$f(x_\tau y_{\overline{\tau}}) = g(\tau;x,y) \approx_{O(\epsilon)} \tau_1 \cdot g_1(x) + \tau_2 \cdot g_2(x) \cdots \tau_n \cdot g_n(x) + g_0(x).$$

By Claim 4.5, $f(x_\tau y_{\overline{\tau}})$ has the approximate form

$$f(x_\tau y_{\overline{\tau}}) \approx_{O(\epsilon)} \sum \tau_i \tilde{g}_i(x_i) + \tilde{g}_0 \tag{10}$$

for some $\tilde{g}_i$ and constant $\tilde{g}_0$. By symmetry, we also have

$$f(x_{\overline{\tau}} y_\tau) \approx_{O(\epsilon)} \sum \overline{\tau}_i \tilde{g}_i(x_i) + \tilde{g}_0 \tag{11}$$

Finally, by (9),

$$f(x) \approx_{12\epsilon} f(x_\tau y_{\overline{\tau}}) + f(x_{\overline{\tau}} y_\tau) - f(y). \tag{12}$$

Plugging (10) and (11) into (12) yields the desired direct sum approximation for f. □

For a function $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$, and subset $T \subseteq [n]$, let $\mathrm{Inf}(T;f)$, denote the influence of the set of variables T on f. That is, $\mathrm{Inf}(T;f) \doteq \Pr_{u,v}[f(u) \neq f(v_T u)]$. Recall that for a pair of disjoint sets (S,T) that do not necessarily partition $[n]$,

$$D_f(S,T;x,y) = f(x) - f(y_S x) - f(y_T x) + f(y_{S\cup T} x).$$

Fact 4.8 (Monotonicity of Influence). *If $S \subseteq S'$ then $\mathrm{Inf}(S;f) \leq \mathrm{Inf}(S';f)$.*

Claim 4.9 (Monotonicity of D_f). *For disjoint sets $S, T \subseteq T'$, $D_f(S,T) \leq D_f(S,T')$.*

Proof. We have

$$D_f(S,T) = \Pr_{x,y}[D_f(S,T;x,y) \neq 0] = \Pr_{x,y}[f(x) - f(y_S x) \neq f(x_S y_T x) - f(y_S y_T x)].$$

For each choice of (x_S, y_S) consider the function $g_{x_S,y_S} : \prod_{i \notin S} \mathcal{D}_i \to \mathcal{G}$ defined as $g_{x_S,y_S}(u) = f(x_S u) - f(y_S u)$. We can rewrite $D_f(S,T)$ as

$$D_f(S,T) = \mathbb{E}_{x_S,y_S} \Pr_{u,v}[f(x_S u) - f(y_S u) \neq f(x_S v_T u) - f(y_S v_T u)] = \mathbb{E}_{x_S,y_S} \mathrm{Inf}(T; g_{x_S,y_S})$$

Thus, by the monotonicity of influence

$$D_f(S,T) = \mathbb{E}_{x_S,y_S} \mathrm{Inf}(T; g_{x_S,y_S}) \leq \mathbb{E}_{x_S,y_S} \mathrm{Inf}(T'; g_{x_S,y_S}) = D_f(S,T').$$ □

Claim 4.10. *For every $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$, $\rho_T(f) \le \rho(f)$.*

Proof. Pick a uniformly random partition (U, V, U', V') of $[n]$. Let $S = U \cup U'$, so $\overline{S} = V \cup V'$. By the monotonicity of D_f and the symmetry of its arguments,

$$\rho_T(f) = \mathbb{E}_{U,V}[D_f(U,V)] \le \mathbb{E}_{S,V}[D_f(S,V)] \le \mathbb{E}_S[D_f(S,\overline{S})] = \rho(f). \qquad \square$$

Proof of Proposition 4.1. Assume $\delta(f) = O(\rho(f))$ for Boolean domain f. By the soundness of T_f (Claim 4.7), $\delta(f) = O(\rho_T(f))$ for every f over arbitrary product domain. By Claim 4.10, $\rho_T(f) = O(\rho(f))$ from where Proposition 4.1 follows. □

5 Testing $\oplus$-Partitionability

Recall that a function $f : \mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$ is *k-$\oplus$-partitionable* if there exists a k-partition $S_1, \dots, S_k$ of $[n]$, and functions $f_1, \dots, f_k$ such that $f(x) = f_1(x_{S_1}) + \ldots + f_k(x_{S_k})$, for all $x \in \mathcal{D}$.

The following claim is an immediate consequence of Theorem 1.1, and allows us to determine whether a function f is $\oplus$-partitionable with respect to a fixed partition $S_1, \dots, S_k$.

Claim 5.1. *Let $(S, \overline{S})$ be a random coarsening of a k-partition $(S_1, \dots, S_k)$ obtained by adding each S_i to S with probability $1/2$. If f is ϵ-far from $\oplus$-partitionable with respect to $S_1, \dots, S_k$, then $D_f(S, \overline{S}; x, y)$ is nonzero with probability $\Omega(\epsilon)$.*

To determine whether a function is k-$\oplus$-partitionable, our testers use the 4-query test D_f to group together variables that cannot occur in different partition components. If the tester finds fewer than k groups, it rejects, otherwise it accepts.

5.1 Adaptive Test for $\oplus$-Partitionability

Our test for k-$\oplus$-partitionability (Algorithm 4) seeks to identify a pair of *contractable* variables s, t that must fall in the same component of a partition. Variables s and t are then contracted and the test is repeated until either fewer than k variables are left (giving a certificate of non-partionability) or no contractable candidates can be found.

A sufficient condition for contractability is that $D_f(\{s\}, \{t\}; x, y)$ is nonzero for some assignment x, y. We start by splitting the variables into k components $S_1, \dots, S_k$ arbitrarily and zero-testing $D_f(S, \bar{S}; x, y)$ for a random coarsening of the components into $S, \bar{S}$. By Claim 5.1, the zero-test fails with probability at least $\Omega(\epsilon)$, where ϵ is the distance between f and the set of functions that are $\oplus$-partitionable with respect to $S_1, \dots, S_k$.

Once such a bipartition $S, \bar{S}$ is identified, s and t can be identified via binary search using Algorithm 3 below. The same idea was used by Blais [Bla09] to identify an influential variable in his junta test. Our t is in fact the influential variable in the function $g(x_{[n]\setminus S}) = f(x) - f(y_S x)$, for fixed x_S, y_S, returned by Blais' test.

```
Algorithm 3: Violating pair adaptive search
  Oracle : $f\colon \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$
  Input  : $(S, T; x, y)$ such that $D_f(S, T; x, y) \neq 0$.
  Output: $(\{s\}, \{t\})$ such that $D_f(\{s\}, \{t\}; x', y') \neq 0$ for some $x', y'$.
  If $|S| = |T| = 1$, output $S, T$.
  If $|T| = 1$, swap $S$ and $T$.
  do
      Split $T$ into two subsets $T'$ and $T''$ of (almost) equal size.
      If $D_f(S, T'; x, y) \neq 0$, recursively run on input $(S, T'; x, y)$.
      Otherwise, recursively run on input $(S, T''; y, x)$.
```

The correctness of Algorithm 3 is based on the following identity.

Claim 5.2. *$D_f(S, T \cup T'; x, y) = D_f(S, T; x, y) + D_f(S, T'; y, x)$ for disjoint sets S, T, T'.*

Proof. Without loss of generality take $S = \{1\}$, $T = \{2\}$, $T' = \{3\}$ and assume there are no other inputs (they are all fixed). By the definition of D_f,

$$\begin{aligned}
D_f(\{1\}, \{2\}; x, y) &= f(x_1 x_2 x_3) + f(y_1 y_2 x_3) - f(x_1 y_2 x_3) - f(y_1 x_2 x_3) \\
D_f(\{1\}, \{3\}; y, x) &= f(y_1 y_2 y_3) + f(x_1 y_2 x_3) - f(y_1 y_2 x_3) - f(x_1 y_2 y_3) \\
-D_f(\{1\}, \{2, 3\}; x, y) &= -f(x_1 x_2 x_3) - f(y_1 y_2 y_3) + f(x_1 y_2 y_3) + f(y_1 x_2 x_3).
\end{aligned}$$

The terms on the right hand side cancel out. □

Lemma 5.3. *Algorithm 3 is correct and has query complexity at most $4(\lceil \log |S| \rceil + \lceil \log |T| \rceil)$.*

Proof. The correctness follows from Claim 5.2 and from the symmetry of D_f in the S, T inputs. As for the query complexity, the algorithm makes four queries (in fact at most two additional queries) in each iteration, and each iteration shrinks one of the original inputs S, T by half. □

In the following algorithm we let $\mathcal{P}(S_1, \ldots, S_k)$ be the distribution on disjoint pairs of sets $(S, \bar{S})$ from Claim 5.1.

```
Algorithm 4: Adaptive tester for $k$-$\oplus$-paritionability.
  Oracle   : $f\colon \mathcal{D}_1 \times \ldots \times \mathcal{D}_n \to \mathcal{G}$
  Input    : Size $k$ of partition
  If $f$ has fewer than $k$ variables, output "not partitionable".
  Otherwise, partition variables arbitrarily into $k$ sets $S_1, \ldots, S_k$.
  repeat
      Choose sets $(S, \bar{S})$ at random from $\mathcal{P}(S_1, \ldots, S_k)$.
      Choose random inputs $x, y$.
  until $D_f(S, T; x, y) \neq 0$;
  Run violating pair adaptive search on input $(S, \bar{S}; x, y)$ to obtain outputs $\{s\}, \{t\}$.
  Contract variables $s$ and $t$ in the oracle and repeat.
```

Proof of Theorem 1.2. We analyze Algorithm 4. Suppose that f is k-$\oplus$-partitionable. By Lemma 5.3, f only contracts variables s, t that are not split by the partition (otherwise $D_f(\{s\}, \{t\}; x', y')$ always vanishes). Therefore f cannot be contracted down to $k-1$ inputs and the tester accepts with probability one.

Now assume f is ϵ-far from partitionable. We will argue that Algorithm 4 outputs "not partitionable" after $O((n-k+1)(\log n + 1/\epsilon))$ queries in expectation by induction on n. Assume $n \geq k$. By Claim 5.1, Loop 6 takes $O(1/\epsilon)$ iterations to complete in expectation, and each iteration costs four queries to f. By Lemma 5.3, line 7 takes another $O(\log n)$ queries. After merging s and t the resulting function on $n-1$ inputs can only be farther from partitionable, so by inductive assumption the expected query complexity $Q(n)$ is at most $Q(n-1)+O(\log n+1/\epsilon)$. This gives the desired bound. By Markov's inequality, Algorithm 4 makes at most twice this number of queries with probability at least half.

The query complexity can be improved slightly to the stated bound $O((n-k)(\log n+1/\epsilon)+1/\epsilon)$ by observing that the violating pair search in line 7 can be bypassed when $n = k$ since a proof of non-partitionability has already been discovered in line 6. □

5.2 Non-adaptive Test for k-$\oplus$-Partitionability

For a pair of disjoint subsets S, T of $[n]$, let $D_f(S,T) \doteq \Pr_{x,y}[D_f(S,T;x,y) \neq 0]$. let G_f denote the graph over the variable set $[n]$, and edge set $\{(i,j) \mid D_f(\{i\},\{j\}) \not\equiv 0\}$. Note that f is k-partitionable into a sum if and only if G_f has k connected components. In order to determine whether f is k-partitionable, our algorithm attempts to determine whether the number of connected components in G_f is at least k.

In each iteration our non-adaptive tester for $\oplus$-partitionability queries all $\binom{n}{2}$ pairs of variables, i.e., for each pair i, j, the algorithm tests whether $D_f(\{i\},\{j\};x,y) \neq 0$ for a random pair of inputs $x, y \in \mathcal{D}$. This is similar to the non-adaptive test in Bogdanov and Wang [BW20], but our analysis improves the number of iterations required. Whereas the tester in [BW20] required $O(nk/\epsilon)$ iterations, we only need $O(k/\epsilon \log^2 n)$. Our savings come from two sources, first the bound in Corollary 5.5 is a factor of k better than the corresponding bound in [BW20], and second we give a tighter analysis for the number of iterations required to find a crossing edge for all k partitions.

Proposition 5.4 (Subadditivity). *For disjoint S, T', T'', $D_f(S, T' \cup T'') \leq D_f(S,T') + D_f(S,T'')$.*

Proof. By Claim 5.2 and a union bound, for a random choice of x and y

$$\Pr[D_f(S, T' \cup T''; x, y) \neq 0] \leq \Pr[D_f(S,T';x,y) \neq 0] + \Pr[D_f(S,T'';y,x) \neq 0]. \qquad \square$$

The following corollary is an immediate from Claim 5.1 and the subadditivity of D_f.

Corollary 5.5. *There is a universal constant C such that if f is ϵ-far from a k-$\oplus$-partitionable function $f_1(x_{S_1}) + \cdots + f_k(x_{S_k})$, then $\sum \Pr[D_f(\{i\},\{j\};x,y) \neq 0] \geq \epsilon/C$, where the sum is over pairs of indices $\{i,j\}$ that belong to different parition components in $(S_1, \ldots, S_k)$, and x, y are chosen independently at random.*

In the following algorithm, let C be the absolute constant in the statement of Corollary 5.5.

Algorithm 5: Non-adaptive tester for k-$\oplus$-partitionability.

```
  Oracle: $f : \mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{G}$
  Input : $k, \epsilon$
1 Initialize $G$ to be the empty graph over vertex set $[n]$.
2 do $\lceil \log n \rceil (4 \lceil C \log(\binom{n}{k-1})) / \epsilon \rceil)$ times
3     foreach pair of distinct inputs $i, j \in [n]$ do
4         Sample independent random points $x, y$ from $\mathcal{D}$.
5         if $D_f(\{i\}, \{j\}; x, y) \neq 0$ then
6             Add edge $(i, j)$ to $G$.
7 Accept if $G$ has at least $k$ connected components.
```

Theorem 5.6. *Algorithm 5 is an $O(n^2 k \log^2(n)/\epsilon)$-query non-adaptive one-sided error tester for k-$\oplus$-partitionability.*

We will need the following fact:

Fact 5.7. *If $p_1, \ldots, p_m$ are probabilities such that $\sum p_i \geq \epsilon$, then $(1 - \prod(1 - p_i)) \geq \epsilon - \epsilon^2/2$.*

Proof of Theorem 5.6. The query complexity follows because each iteration of the outer loop requires $O(n^2)$ queries, and $\log \binom{n}{k-1} = \Theta(k \log n)$.

The tester has perfect completeness because the connected components of G are always contained in the partition components of f.

Now, suppose f is ϵ-far from being k-partitionable. We view the algorithm as being made up of $\lceil \log n \rceil$ rounds, where each round is made up of $4\lceil C \log \binom{n}{k-1}/\epsilon \rceil$ iterations of the outer loop. For $\ell \in [\lceil \log n \rceil]$, let $\mathcal{C}_\ell$ denote the set of connected components of G at the beginning of the ℓ-th round. We claim the the following invariant holds:

Claim 5.8. *For each $\ell \in [\lceil \log n \rceil]$, $|\mathcal{C}_\ell \cap \mathcal{C}_{\ell+1}| \leq k - 2$ with probability at least $(1 - 1/n)$.*

Assuming the claim for the moment, we prove soundness of the tester. By the claim, for each ℓ, $|\mathcal{C}_{\ell+1}| \leq (|\mathcal{C}_\ell| - (k-2))/2 + (k-2)$, with probability at least $(1 - 1/n)$. Thus, by a union bound, $|\mathcal{C}_{\log n + 1}| \leq k - 1$ with probability at least $(1 - \log n/n)$. □

Proof of Claim 5.8. Fix ℓ. Fix a k-partition, $\mathcal{P}$ induced by $k-1$ elements of $\mathcal{C}_\ell$, i.e., $k-1$ of the partition components in $\mathcal{P}$ are elements of $\mathcal{C}_\ell$.

Consider an iteration within round ℓ. The probability that the test at Line 5 evaluates to true for some $i, j \in [n]$ is $\Pr[D_f(\{i\}, \{j\}; x, y) \neq 0]$. By of Corollary 5.5 and Fact 5.7, the probability that an edge crossing the partition $\mathcal{P}$ is picked in this iteration is at least $\epsilon/2C$. Thus, the probability that an edge crossing $\mathcal{P}$ is picked in round ℓ is at least $(1 - e^{-2\log \binom{n}{k-1}})$.

There are at most $\binom{n}{k-1}$ choices for the partition $\mathcal{P}$. By a union bound, an edge crossing $\mathcal{P}$ is picked, for every choice of $\mathcal{P}$ with probability at least $(1 - e^{-\log \binom{n}{k-1}}) \geq (1 - 1/n)$. □

5.3 Non-adaptive Lower Bound for $\oplus$-Partitionability

In [Bla08], Blais showed that any non-adaptive algorithm for testing j-juntas requires $\Omega\left(\frac{j/\epsilon}{\log j/\epsilon}\right)$ queries. Based on ideas of Chockler and Gutfreund [CG04], Blais constructed two distributions

D_{yes} and D_{no} over functions $f : \{0,1\}^n \to \{0,1\}$ that are indistinguishable by any set of $o\left(\frac{j/\epsilon}{\log j/\epsilon}\right)$ non-adaptive queries such that functions in D_{yes} are always j-juntas and functions in D_{no} are are ϵ-far from j-juntas with high probability. Servedio et al. [STW15] improved the analysis of Blais and showed that any set of queries that distinguish between the distributions of Blais require $\Omega\left(\frac{j \log j}{\epsilon^c \log\log j/\epsilon^c}\right)$ queries, where c is an absolute constant less than 1.

Since any $(n-k+1)$-junta is k-partitionable into a sum, the lower bound of Servedio et al. for $j = n-k+1$, also gives a lower bound for k-partitionability, provided the elements in D_{no} for are ϵ-far from k-partitionable with high probability. We show that this is indeed the case.

D_{yes} and D_{no} are defined as follows. Let $j = (n-k+1)$. A sample f from D_{yes} is generated by setting $f(x) = g(x_{[j]})$, for a random function $g : \{0,1\}^j \to \{0,1\}$, such that for each $x \in \{0,1\}^m$, $g(x)$ is chosen independently at random with $\Pr[g(x) = 1] = \epsilon$. D_{no} is defined identically except using a random function g with domain $\{0,1\}^{j+1}$.

Claim 5.9. *For $20\log(n-k+2)/2^{(n-k+2)} \leq \epsilon \leq 1/10$, a random sample from D_{no} is $\epsilon/40$ far from $\oplus$-partitionable into sums with probability at least $(1 - 1/n - k + 1)$.*

We use the following notation in the proof: We write $\delta_f(\{i\},\{j\})$ for the distance between f and the closest function f' that can be written as $f'_1 + f'_2$, such that f'_1 doesn't depend on $\{j\}$ and f'_2 doesn't depend on $\{i\}$. We note that both f'_1, f'_2 could depend on the remaining variables.

Proof. Let $m = n-k+2$. Let $g : \{0,1\}^m \to \{0,1\}$ be a random function such that for every $x \in \{0,1\}^m$, $g(x)$ is chosen independently at random with $\Pr[g(x) = 1] = \epsilon$. It is sufficient to show that g is far from $\oplus$-partitionable into two or more components.

For $i, j \in [m]$, let t_{ij} denote the number of strings $x \in \{0,1\}^m$ such that $g(x) + g(x^i) + g(x^j) + g(x^{ij}) \neq 0$, where x^i is x with the i-th bit flipped, and x^{ij} is x with the i and j-th bits flipped. The distance between g and the closest function f with $\delta_f(\{i\},\{j\}) = 0$ is $t_{ij}/2^{m-2}$.

The expected value of t_{ij} is at least $9\epsilon 2^{m-2}/10$. By Chernoff's bound,

$$\Pr[t_{ij} < \epsilon 2^{m-2}/10] \leq e^{-9\epsilon 2^{m-2}/10 \cdot (1-1/9)^2/2} \leq e^{-(1/10)\epsilon 2^m} \leq 1/m^2.$$

Taking a union bound over the pairs $(1,2), (2,3), \ldots, (m-1,m)$, we get that with probability at least $(1 - 1/m)$, for every partition $(S, \bar{S})$, g is $\epsilon/10$-far from being partitionable with respect to $(S, \bar{S})$. □

By the lower bound in [STW15], we get the following proposition.

Proposition 5.10. *There exists an absolute $c < 1$ such that for $30\log(n-k+2)/2^{(n-k+2)} \leq \epsilon \leq 1/10$, any ϵ-tester for k-$\oplus$-partitionability must make $\Omega\left(\frac{(n-k+1)\log(n-k+1)}{\epsilon^c \log\log(n-k+1)/\epsilon^c}\right)$ queries.*

6 Testing $\otimes$-Partitionability

Recall that a function $f : \mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{R}^k$ is *k-$\otimes$-partitionable* if there exists a k-partition $S_1, \ldots, S_k$ of $[n]$, and functions $f_1, \ldots, f_k$ such that $f(x) = (f_1(x_{S_1}), \ldots, f_k(x_{S_k}))$ for all $x \in \mathcal{D}$.

In this section we present a $O((n/\epsilon)\log^2(n/\epsilon))$-query non-adaptive one-sided error test for $\otimes$-partitionability (see Theorem 1.3). We begin with an overview of the construction.

First, some notation. For a function $f : \mathcal{D} \to \mathcal{R}^k$ and a subset $T \subseteq [k]$ we write f_T to refer to the function obtained by projecting the output of f onto the coordinates in T. We often write x_i instead of $x_{\{i\}}$ and f_j instead of $f_{\{j\}}$.

For simplicity, suppose that $k = 2$. Then f is 2-$\otimes$-partitionable if and only if every variable has non-zero influence on at most one of the two coordinates of f. So our task boils down to determining whether there is a variable that is influential in both coordinates of the output of f. The key observation that allows us to find such a coordinate with a small number of queries is that if f is ϵ-far from $\otimes$-partitionable, then

$$\sum_{i\in[n]} \min(\mathrm{Inf}(i; f_1), \mathrm{Inf}(i; f_2)) \geq \epsilon. \tag{13}$$

Therefore, if f is ϵ-far from $\otimes$-partitionable, there must be a coordinate that has influence at least ϵ/n in both coordinates. This immediately suggests an $O(n^2/\epsilon)$ query test: for each variable use $O(n/\epsilon)$ queries to determine whether it is influential in both coordinates.

We obtain an improvement in the query complexity by exploiting a trade-off between the number of samples, $i \in [n]$, required to find a variable that is influential in both coordinates, and the number of samples required to certify that a variable is indeed influential in both coordinates.

Given subsets $S_1, \dots, S_k$ of $[n]$, let $\Delta_f(S_1, \dots, S_k)$ be the distance from $f = (f_1, \dots, f_k)$ to the closest function $g = (g_1, \dots, g_k)$ in which g_j does not depend on the inputs in S_j, and define the *influence* of $(S_1, \dots, S_k)$ on f as

$$\mathrm{Inf}(S_1, \dots, S_k; f) = \Pr[f_j(x) \neq f_j(y_{S_j} x) \text{ for some } j],$$

where x, y is an independent pair of inputs.

Proposition 6.1. $\mathrm{Inf}(S_1, \dots, S_k; f) \leq \sum_{i=1}^n \mathrm{Inf}(i; f_{J(i)})$, *where* $J(i)$ *is the set of output coordinates* $j \in [k]$ *for which* S_j *contains* i.

Proof. Let E be the event "$f_j(x) \neq f_j(y_{S_j} x)$ for some j". Let h^i be the hybrid input in which $h^i_t = y_t$ for $t \leq i$ and x_t for $t > i$. Then $h^0 = x$ and $h^n = y$. If the event "$f_j(x) \neq f_j(y_{S_j} x)$" occurs, then one of the events "$f_j(h^{i-1}_{S_j} x) \neq f_j(h^i_{S_j} x)$" must occur for some i between 1 and n. By the union bound, $\Pr(E) \leq \sum_{i=1}^n \Pr(E_i)$, where E_i is the event "$f_j(h^{i-1}_{S_j} x) \neq f_j(h^i_{S_j} x)$ for some j." The inputs $h^{i-1}_{S_j} x$ and $h^i_{S_j} x$ are identical unless $i \in S_j$, in which case they differ only in the i-th coordinate where they are independent. Therefore

$$\Pr(E_i) = \Pr[f_j(x) \neq f_j(x^i) \text{ for some } j \in J(i)],$$

where x^i is x with its i-th input resampled independently. The right hand side is precisely the influence of i in $f_{J(i)}$. □

Claim 6.2. $\Delta_f(S_1, \dots, S_k) \leq \mathrm{Inf}(S_1, \dots, S_k; f)$.

Proof. By averaging, there must exist an assignment a to y such that

$$\mathrm{Inf}(S_1, \dots, S_k; f) \geq \Pr[f_j(x) \neq f_j(a_{S_j} x) \text{ for some } j].$$

Define $g_j(x) = f_j(a_{S_j} x)$ (on all inputs). Then g_j does not depend on the inputs in S_j, so

$$\Delta_f(S_1, \dots, S_k) \leq \Pr[f_j(x) \neq g_j(x) \text{ for some } j] = \Pr[f_j(x) \neq f_j(a_{S_j} x) \text{ for some } j] \leq \Pr(E).$$

□

Claim 6.3. *Let $k \geq 2$ and $f(x_i) = (f_1(x_i), \dots, f_k(x_i))$ be a possibly randomized univariate function. Let d an the output coordinate that maximizes $\mathrm{Inf}(i; f_d)$. There exists a partition $(P, \overline{P})$ of the output coordinates such that $\mathrm{Inf}(i; f_P)$ and $\mathrm{Inf}(i; f_{\overline{P}})$ are both at least $\mathrm{Inf}(i; f_{[n]\setminus\{d\}})/3$.*

Proof. Let I_j be the event $f_j(x) \neq f_j(y)$ for random independent x and y. Then $\mathrm{Inf}(i; f_T) = \Pr(\cup_{j \in T} I_j)$. Let $\delta_i = \Pr(\cup_{j \neq d} I_j)$. If $\Pr(I_d) \geq \delta_i/3$ then the partition $(\{d\}, [n] \setminus \{d\})$ satisfies the conclusion. Otherwise, $\Pr(I_j) \leq \delta_i/3$ for all j. Then some partition of type $(I_1 \cup \dots \cup I_j, I_{j+1} \cup \dots \cup I_k)$ works: If j is the first set for which $\Pr(I_1 \cup \dots \cup I_j)$ exceeds $\delta_i/3$, then

$$\Pr(I_1 \cup \dots \cup I_j) \leq \Pr(I_1 \cup \dots \cup I_{j-1}) + \Pr(I_j) \leq 2\delta_i/3.$$

Since

$$\Pr(I_1 \cup \dots \cup I_j) + \Pr(I_{j+1} \cup \dots \cup I_k) \geq \Pr(\cup_j I_j) \geq \delta_i,$$

the event $I_{j+1} \cup \dots \cup I_k$ also has probability at least $\delta_i/3$. □

Lemma 6.4. *If f is δ-far from $\otimes$-partitionable then there exist partitions $(P(1), \overline{P(1)})$, $\dots$, $(P(n), \overline{P(n)})$ of $[k]$ such that*

$$\sum_{i=1}^{n} \min\big\{\mathrm{Inf}(i; f_{P(i)}), \mathrm{Inf}(i; f_{\overline{P(i)}})\big\} \geq \frac{\delta}{3}.$$

Proof. Let $j^*(i)$ be the maximizer of $\mathrm{Inf}(i; f_j)$ (breaking ties arbitrarily), and S_j be the set of all i such that $j^*(i) \neq j$. Then $J(i) = \{j \colon i \in S_j\} = [n] \setminus \{j^*(i)\}$. By Proposition 6.1 and Claim 6.2, $\delta \leq \Delta_f(S_1, \dots, S_k) \leq \sum \mathrm{Inf}(i; f_{[n]\setminus\{j^*(i)\}})$. By Claim 6.3 applied to f as a function of x_i only (randomized over the other inputs), $\mathrm{Inf}(i; f_{[n]\setminus\{j^*(i)\}})/3 \leq \min\big\{\mathrm{Inf}(i; f_{P(i)}), \mathrm{Inf}(i; f_{\overline{P(i)}})\big\}$. □

Algorithm 6: Non-adaptive tester for $\otimes$-partitionability.

```
  Oracle: $f : \mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{R}^k$
  Input : Proximity parameter $\epsilon$
1 foreach $r \in \{0, \dots, \lceil \log(3n/\epsilon) \rceil\}$ do
2     Let $S \subseteq [n]$ be a set of $3 \cdot \lceil \frac{6n \log(3n/\epsilon)}{2^r \epsilon} \rceil$ indices sampled uniformly at random from $[n]$.
3     foreach $i \in S$ do
4         Sample $3 \cdot 2^{r+1}$ independent pairs of inputs from $\mathcal{D}$.
5         if $\exists$ samples $(x, y), (x', y')$, and $j \neq j' \in [k]$ such that
           $f_j(x) \neq f_j(y_{\{i\}}x)$ and $f_{j'}(x') \neq f_{j'}(y'_{\{i\}}x')$ then
6             Reject.
7 Accept.
```

Proof of Theorem 1.3. We show that Algorithm 6 satisfies the statement of the theorem. In each iteration of the outer loop, $O(n/\epsilon \log(n/\epsilon))$ queries are made to f. Thus, in total the algorithm makes $O((n/\epsilon) \log^2(n/\epsilon))$ queries.

The test has perfect completeness because the condition on Line 5 is never triggered if f is a direct product.

We now argue soundness. If f is ϵ-far from being a direct product, then by Lemma 6.4, for every $i \in [n]$ there exist partitions $(P(i), \overline{P(i)})$ such that

$$\sum_{i \in [n]} M_i \geq \epsilon/3, \tag{14}$$

where $M_i = \min\{\mathrm{Inf}(i; f_{P(i)}), \mathrm{Inf}(i; f_{\overline{P(i)}})\}$.

For $r \in \{0, \dots, \lceil \log(3n/\epsilon) \rceil\}$, let A_r denote the set $\{i \mid M_i \in [1/2^r, 1/2^{r+1})\}$. By (14) and an averaging argument, we know that there exists an ℓ such that $|A_\ell| \geq \lceil \frac{2^\ell \epsilon}{6 \log(3n/\epsilon)} \rceil$. For such an ℓ, we show that the probability that the algorithm rejects in the ℓ-th iteration is at least $2/3$.

Consider the ℓ-th iteration of the outer loop. The probability that no index in A_ℓ is picked at Line 2 is at most $(1 - \frac{|A_\ell|}{n})^{3 \cdot \frac{n}{|A_\ell|}} \leq 1/e^3$.

In an iteration of the inner loop corresponding to an index $i \in A_\ell$, the probability that either $f_{P(i)}(x) = f_{P(i)}(y)$ for all sampled pairs (x, y), or $f_{\overline{P(i)}}(x) = f_{\overline{P(i)}}(y)$ for all sampled pairs (x, y) is at most $2 \cdot (1 - 1/2^{\ell+1})^{3 \cdot 2^{\ell+1}} \leq 2/e^3$. This tells us that the probability that the algorithm rejects in the ℓ-th iteration of the outer loop conditioned on $A_\ell \cap S \neq \emptyset$ is at least $(1 - 2/e^3)$. Since $A_\ell \cap S$ is empty with probability at most $1/e^3$, the probability that the algorithm rejects in the ℓ-th iteration is at least $(1 - 3/e^3) \geq 2/3$. □

6.1 Lower Bounds for Testing ⊗-partitionability

We show that any 1/2-tester for ⊗-partitionability requires at least $\Omega(n)$ queries, and that any non-adaptive ϵ-tester for direct products requires at least $O\left(\frac{n}{\epsilon \log(1/\epsilon)}\right)$ queries.

Lower bound for adaptive testers: We obtain a lower bound for adaptive testers by a reduction from set disjointness [BBM12]. In the set disjointness problem, Alice is given a set $S \subseteq [n]$ and Bob is given a set $T \subseteq [n]$, and the goal is to determine whether S and T are disjoint. It is well known that any bounded-error two-party communication protocol for set disjointness, in the shared randomness model, requires at least $\Omega(n)$ bits of communication [SK87, Raz92]. Our query lower bound is obtained by showing that if there is a q-query tester for ⊗-partitionability then there is a protocol for set disjointness of complexity $O(q)$.

Claim 6.5. *Let $S, T \subseteq [n]$ be sets with non-empty intersection. Then, the function $f : \{0,1\}^n \to \{0,1\}^2$ defined as $f(x) = (\oplus_{i \in S} x_i, \oplus_{j \in T} x_j)$ is 1/2-far from ⊗-partitionable.*

Proof. Let $(F, G) : \{0,1\}^n \to \{0,1\}^2$ be a ⊗-partitionable function. Without loss of generality, assume the support of F differs from S. Then, F and $\oplus_{i \in S} x_i$ differ on half the inputs. □

Proposition 6.6. *Determining whether a function $f : \{0,1\}^n \to \{0,1\}^2$ is a direct product, or 1/2-far, with respect to the uniform distribution, from direct products, requires at least $\Omega(n)$ queries.*

Proof. Let $S \subseteq [n]$ be Alice's input, and $T \subseteq [n]$ be Bob's input. Alice forms the function $f \doteq \oplus_{i \in S} x_i$ and Bob forms the function $g \doteq \oplus_{i \in T} x_i$. If S and T are disjoint, then (f, g) is a direct product. By Claim 6.5, if S and T are not disjoint, (f, g) is at least 1/2-far from a direct product.

Alice and Bob can simulate any tester for direct products as follows: For each query x, Alice determines $f(x)$ and sends it to Bob, and Bob determines $g(x)$ and sends it to Alice. Thus, both Alice and Bob can determine the value of (f, g) on any query with just two bits of communication. It follows that any tester for direct product requires $\Omega(n)$ queries. □

Lower bound for non-adaptive testers: We describe two distributions D_{yes} and D_{no} over $\otimes$-partitionable, and ϵ-far from $\otimes$ paritionable functions, respectively, and argue that any non-adaptive tester must make $\Omega\left(\frac{n}{\epsilon \log 1/\epsilon}\right)$ queries to distinguish between D_{yes} and D_{no} with constant probability. Our distributions are based on the distributions of Blais [Bla08] described in Section 5.3.

We say two strings $x, y \in \{0,1\}^n$ are i-twins if they only differ at the i-th coordinate. For $x \in \{0,1\}^n$, let x^i denote the i-twin of x.

Let $D_{yes}^{(i)}$ be the distribution over functions $F = (F_1, F_2) : \{0,1\}^n \to \{0,1\}^2$, generated by picking $F_1 : \{0,1\}^{[n]\setminus\{i\}} \to \{0,1\}$ so that for every $x \in \{0,1\}^{[n]\setminus\{i\}}$, $F_1(x)$ is set independently at random with $\Pr[F_1(x) = 1] = \epsilon$, and by fixing $F_2(x)$ to be x_i. Let D_{yes} be the uniform mixture of $D_{yes}^{(1)}, \ldots, D_{yes}^{(n)}$.

Let $D_{no}^{(i)}$ be the distribution over functions $G = (G_1, G_2) : \{0,1\}^n \to \{0,1\}^2$, generated by picking $G_1 : \{0,1\}^n \to \{0,1\}$ so that so that for every $x \in \{0,1\}^n$, $G_1(x)$ is set independently at random with $\Pr[G_1(x) = 1] = \epsilon$, and fixing $G_2(x)$ to be x_i. Let D_{no} be the uniform mixture of $D_{no}^{(1)}, \ldots, D_{no}^{(n)}$.

Claim 6.7. *For $100/2^n \leq \epsilon \leq 1/2$, a random sample from D_{no} is $\epsilon/10$-far from $\otimes$-partitionable with probability at least $15/16$.*

Proof. It suffices to prove the statement for a function sampled from from $D_{no}^{(i)}$, for $i \in [n]$.

Let G be a function sampled from $D_{no}^{(i)}$. Let (f_1, f_2) be $\otimes$-partitionable. If f_2 does not depend on x_i, then G is $1/2$-far from (f_1, f_2) because x_i and f_2 disagree on half the inputs.

In the case that f_2 depends on x_i, it suffices to show that G_1 is $\epsilon/10$-far from any function that does not depend on x_i. Let t denote the number of i-twins (x, x^i) such that $G_1(x) \neq G_1(x^i)$. The distance between G_1 and the closest function that does not depend on x_i is $t/2^n$. The expected value of t is $\epsilon(1-\epsilon)2^n \geq (\epsilon/2)2^n$. By Chernoff's bound,

$$\Pr[t < \epsilon 2^n/10] \leq e^{-\epsilon 2^{n-1}(1-1/5)^2/2} = e^{-\epsilon 2^n/25}.$$

□

For a set of queries Q, let $R_{yes}^{(i)}(Q)$ be the distribution of the values of a function sampled from $D_{yes}^{(i)}$ on Q, and let $R_{no}^{(i)}(Q)$ be the distribution of values of a function sampled from $D_{no}^{(i)}$ on Q. For fixed i, the value of x_i in the replies to the query set has no effect on the statistical distance between the distributions $R_{yes}^{(i)}(Q)$ and $R_{no}^{(i)}(Q)$. Similar to Lemma 4.3 and in [Bla08], we have the following fact.

Let $d_{ST}(D_1, D_2)$ denote the statistical distance between distributions D_1 and D_2.

Claim 6.8. *There exists a constant $c > 0$ such that for any set of queries Q with t i-twins,*

$$d_{ST}(R_{yes}^{(i)}(Q), R_{no}^{(i)}(Q)) \leq ct\epsilon.$$

By a triangle inequality we have that for any set of queries Q that contain at most t i-twins in all coordinates $d_{ST}(R_{yes}(Q), R_{no}(Q)) = O(t\epsilon)$, where R_{yes}, and R_{no} are the distributions of function values on the query set Q, for functions sampled from D_{yes} and D_{no} respectively.

We now employ an edge isoperimetric inequality Servedio et al.:

Claim 6.9 ([STW15])**.** *Any set $Q \subseteq \{0,1\}^n$ that contains t i-twins for $\Omega(n)$ directions i must have size at least $\Omega(tn/\log t)$*

Putting everything together and using Yao's Minmax Principle we obtain the following lower bound:

Proposition 6.10. *For $100/2^n \le \epsilon \le 1/2$, determining whether a function $f : \{0,1\}^n \to \{0,1\}^2$ is $\otimes$-partitionable or ϵ-far from $\otimes$-partitionable requires $\Omega\left(\frac{n}{\epsilon \log(1/\epsilon)}\right)$ non-adaptive queries.*

Acknowledgement

We thank Chandra Nair and Yan Nan Wang for valuable discussions on the hypercontractivity of information channels.

References

[AG76] Rudolf Ahlswede and Peter Gacs. Spreading of sets in product spaces and hypercontraction of the markov operator. *Ann. Probab.*, 4(6):925–939, 12 1976.

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Comput. Complex.*, 21(2):311–358, 2012.

[Bla08] Eric Blais. Improved bounds for testing juntas. In Ashish Goel, Klaus Jansen, José D. P. Rolim, and Ronitt Rubinfeld, editors, *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, page 317–330, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[Bla09] Eric Blais. Testing juntas nearly optimally. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 151–158, 2009.

[BLR90] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing*, pages 73–83, New York, NY, USA, 1990. Association for Computing Machinery.

[Bon70] Aline Bonami. Étude des coefficients de Fourier des fonctions de $l^p(g)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970.

[BSSVW03] Eli Ben-Sasson, Madhu Sudan, Salil Vadhan, and Avi Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 612–621, New York, NY, USA, 2003. Association for Computing Machinery.

[BW20] Andrej Bogdanov and Baoxiang Wang. Learning and testing variable partitions. In Thomas Vidick, editor, *11th Innovations in Theoretical Computer Science Conference, ITCS 2020, January 12-14, 2020, Seattle, Washington, USA*, volume 151 of *LIPIcs*, pages 37:1–37:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[CG04] Hana Chockler and Dan Gutfreund. A lower bound for testing juntas. *Inf. Process. Lett.*, 90(6):301–305, 2004.

[DD19] Yotam Dikstein and Irit Dinur. Agreement testing theorems on layered set systems. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1495–1524. IEEE Computer Society, 2019.

[DG19] Irit Dinur and Konstantin Golubev. Direct sum testing: The general case. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA*, pages 40:1–40:11, 2019.

[DS14] Irit Dinur and David Steurer. Direct product testing. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 188–196. IEEE Computer Society, 2014.

[NW16] Chandra Nair and Yan Nan Wang. Evaluating hypercontractivity parameters using information measures. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 570–574, 2016.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, USA, 2014.

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[SK87] Georg Schnitger and Bala Kalyanasundaram. The probabilistic communication complexity of set intersection. In *Proceedings of the Second Annual Conference on Structure in Complexity Theory, Cornell University, Ithaca, New York, USA, June 16-19, 1987*. IEEE Computer Society, 1987.

[STW15] Rocco A. Servedio, Li-Yang Tan, and John Wright. Adaptivity Helps for Testing Juntas. In David Zuckerman, editor, *30th Conference on Computational Complexity (CCC 2015)*, volume 33 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 264–279, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[SW06] Amir Shpilka and Avi Wigderson. Derandomizing homomorphism testing in general groups. *SIAM J. Comput.*, 36(4):1215–1230, 2006.